Stress and accent. Prominence relations in Southern Standard British English

Jensen, Christian

Publication date:
2004

Document version
Publisher's PDF, also known as Version of record

Stress and Accent

Prominence relations in Southern Standard British English

PhD thesis
Submitted to the University of Copenhagen
February 2004

Typeset in groff, 12pt Legacy

A webpage accompanies this thesis.
It can be found at the following two locations:
http://www.cphling.dk/pers/chrjen/thesis/
http://www.fonetik.dk/thesis/

Abstract

This thesis examines the (perceived) prominence relations (stress and accent levels) in Southern Standard British English. In line with previous descriptions it was found that stressed words adjacent to utterance boundaries, or major phrase boundaries, are perceived as more prominent than stressed words in intermediate positions, but contrary to many descriptions in the traditional British school of intonation analysis it was not found that the final stressed word (the so-called nucleus) is generally more prominent than all other words in the phrase. Only 15-20% of the neutral, context-free utterances had a clearly more prominent final item.

In utterances where one word is emphasised due to some type of narrow focus there is both a local effect on the emphasised item, resulting in significantly higher perceived prominence on this word, and a global effect on surrounding words, which are perceived as less prominent; in other words a combined foregrounding and backgrounding of items inside and outside of the focus domain, respectively.

The backgrounding effect is largest in post-focal position, where the prominence level of all stressed words is reduced, regardless of their distance from the focal accent. In pre-focal position the reduction in prominence level is inversely proportional to the distance from the focal accent: immediately adjacent items are reduced the most.

The relevance of these observations was demonstrated in an experiment which examined the relation between perceived prominence and perceived information structure. As expected, listeners perceived utterances in which one item was particularly prominent as responses to questions about this single constituent – they heard the item as being in (narrow) focus. This was true of intended neutral and intended focused utterances alike and in all positions in the utterance, even when the most prominent item was in utterance final position, that is, the default location of the nucleus.

The relative reduction of non-focal items contributed to the perception of focus, and the results suggested that post-focal reduction is more important than pre-focal reduction, in accordance with the results of the prominence perception experiments.

A brief account of the acoustic parameters F_0 and duration is presented. The variation in F_0 mirrored the perceived prominence in a fairly direct way in both neutral utterances and in pre-focal, focal and post-focal position: F_0 movements were almost absent in post-focal position but were reduced in inverse proportion to the distance from the focal accent in pre-focal position. The duration data indicated a larger pre-focal shortening effect and only a very modest post-focal effect. The two acoustic parameters thus seem to operate differently depending on their position relative to the focal accent.

Dansk resumé (Danish summary)

Denne afhandling undersøger de opfattede prominensforhold (eller opfattet grad af *tryk*) i Southern Standard British English. Resultaterne af en række lytteforsøg viste, i lighed med tidligere beskrivelser, at trykstærke ord i umiddelbar nærhed af ytringsgrænser, eller større frasegrænser, opfattes som mere prominente end trykstærke ord i mellemliggende positioner, men i modsætning til mange beskrivelser i den traditionelle britiske intonationsskole kunne det ikke demonstreres at et frasefinalt trykstærkt ord (den såkaldte 'nucleus') generelt set er mere prominent end alle andre ord i frasen. Kun 15–20% af de neutrale, kontekstfri ytringer havde et klart mere prominent finalt ord.

I ytringer hvor et ord er fremhævet for at signalere snæver fokus (semantisk fokus eller kontrastfokus), er der både en lokal effekt på det fremhævede ord, hvilket ses ved at dette ord opfattes som betydeligt mere prominent, og en global effekt på omkringliggende ord, som opfattes som mindre prominente. Der er med andre ord tale om en kombineret effekt af at fremhæve ord indenfor fokusdomænet og nedtone ord udenfor fokusdomænet.

Effekten af at nedtone ord udenfor fokus er størst postfokalt, hvor prominensniveauet på alle trykstærke ord reduceres, uafhængigt af deres afstand til den fokale accent (det fokuserede eller kontrasterede ord). Præfokalt er reduktionen i prominens omvendt proportional med afstanden til den fokale accent: umiddelbart tilstødende ord reduceres mest. Relevansen af disse observationer blev demonstreret i et eksperiment som undersøgte forholdet mellem opfattet prominens og opfattelsen af informationsstruktur. Som ventet opfattede lytterne ytringer hvor et ord var særligt prominent som svar på et spørgsmål om netop dette element – de hørte dette ord som fokuseret. Dette var tilfældet både med ytringer som var ment som neutrale fra talerens side, og med ytringer som var ment som fokuserede, og det gjaldt alle positioner i ytringen – selv når det mest prominente ord var i ytringsfinal position, dvs. den forventede placering af 'nucleus'.

Reduktion af ikke-fokale ord medvirkede til opfattelsen af fokus, og resultaterne antydede at postfokal reduktion er mere væsentlig end præfokal reduktion, hvilket er i overensstemmelse med resultaterne af undersøgelserne af opfattet prominens.

De akustisker parametre F_0 og varighed præsenteres kort. Variationen i F_0 afspejler den opfattede prominens på en forholdsvis direkte måde i både neutrale ytringer og i preæfokal, fokal og postfokal position: F_0-bevægelser var stort set fraværende postfokalt, men i præfokal position var de reduceret omvendt proportionalt med afstanden til det fremhævede ord. Varigheddataene pegede på en større præfokal forkortelse og kun en beskeden postfokal effekt. Dermed synes de to akustiske parametre at virke forskelligt afhængigt af deres position i forhold til den fokale accent.

Acknowledgements

This project was supported by a grant from the Danish Research Council for the Humanities and was carried out at the Department of General and Applied Linguistics, University of Copenhagen; I greatly appreciate both the financial support of the Research Council and the hospitality of the department. I would like to thank all who have helped me in one way or another over the years, and particularly everybody who acted as a speaker or listener in my experiments.

Contents

Figures

Tables

Introduction

Scope and purpose of the investigation

The observable variations in prominence which characterise languages such as English and Danish have been studied for many years and from many different angles. Most attention, especially in earlier accounts, has probably been paid to the lexical aspect of stress: that every word has (at least) one syllable which is normally more prominent than the others, how this affects the rhythmical aspects of an utterance, and how rhythmical considerations may in some cases shift the perception of the location of the prominent, or stressed, syllable. This applies in particular to the relatively small set of words which are distinguished solely or mainly by the location of the stressed syllable, for example *ímport* (noun) versus *impórt* (verb), which have consequently been used in many experiments on stress. Another perspective which has been studied is how placing additional prominence on certain syllables, words or phrases affects our interpretation of an utterance. This was previously often referred to as *emphasis* but is now usually known as *focus*, which shows the interrelation with semantics and pragmatics.

Despite the attention this topic has received over time, relatively little is known about the manifestation of prominence in different positions in English utterances. How is the relative strength or prominence of words or syllables perceived by listeners, and how does their perception relate to the acoustic properties of the stressed items? And how do the variations in stress or prominence level correspond with our perception of information structure in the utterance?

Most theories about English intonation provide an account of the expected levels of stress or accent in an utterance, whether this is seen as determined by the metrical structure of the constituents of an utterance, as in autosegmental-metrical descriptions (Pierrehumbert 1980, Ladd 1996), or by the internal structure of intonation units within the utterance, as in the British school of intonation analysis (O'Connor and Arnold 1973, Cruttenden 1997). The latter framework is particularly interesting from a didactic perspective, since it is by far the most influential school of intonation in relation to the teaching of English as a foreign language in Denmark. The description of intonation in Southern Standard British English in this tradition indicates many fundamental differences between Danish and English intonation, including the expected prominence levels, or stress and accent levels, in certain types of utterance. Lexical (or content) words are normally expected to be stressed in both languages, but according to the British tradition they may have one of three different stress/accent levels depending on their position in the utterance, even in neutral, context-free utterances. In Danish all stressed words are normally assumed to be

equally prominent in this type of utterance. One particular difference concerns the presence of one item which is more prominent than all other stressed items, namely the *nucleus*, which most descriptions consider to be an obligatory element of all English utterances. Standard Danish, however, does not have an obligatory nucleus, or sentence accent.

This thesis examines some of these putative differences between English and Danish through empirical, experimental investigations of prominence relations in Southern Standard British English. Two issues in particular are considered: (1) what are the systematic variations in prominence level in different positions in the utterance, and (2) how does giving emphasis to, or focusing, a specific (single) item affect the prominence of this and surrounding items? The focal point of the investigation is the perceptual aspect – how prominence levels are perceived by listeners. Data have also been collected on the acoustic manifestation of prominence; the main trends are presented separately and serve as a background for the analyses of perceived prominence.

At the outset of this project the acoustic manifestation was intended to be the main consideration, but as the work progressed, the results from the perceptual experiments prompted further investigation into this area. The detailed analyses of the acoustic properties of stress and accent ended up being beyond the scope of this thesis and have had to be deferred until a later time. Many of the preliminaries to such analyses have already been done, however, and the results of the perceptual experiments have been vital in suggesting the proper direction which the present work should take.

The original purpose of doing acoustic analyses has had a considerable influence on many facets of this project, from the choice of experimental material and procedures to the theoretical starting point for the description of stress, accent and prominence. While care has been taken to present the results of the project in the light of the new, current perspective, the original (and future) orientation towards acoustic analyses remains evident to some degree throughout the thesis.

The structure of the thesis

The theoretical background and previous research on stress, accent and prominence is outlined in Chapter 1. Chapter 2 describes the collection of the data material used in all subsequent experiments, and at the end of the chapter the major trends in variation of the acoustic features F_0 and duration are presented. Chapter 3 is an account of the experimental method used in the three listening experiments presented in Chapters 4, 5 and 6 (Tests 1–3), and also addresses the possible effect of experience and background on listeners' perception of prominence levels. Finally, Chapter 7 examines the association between perceived prominence levels and perceived information structure (Test 4).

CHAPTER 1

Stress, accent and prominence

1.1 Introduction

> What many phoneticians and linguists have called *stress*, and what most laymen readily understand under this term, refers to nothing more than the fact that in a succession of spoken syllables or words some will be perceived as more salient or prominent than others.
> (Couper-Kuhlen 1986: 19)

The above quotation provides a simple definition of stress as the relative prominence of syllables or words – a definition which in general agrees well with my understanding of this phenomenon. However, stating that stress refers to 'nothing more' than relative prominence is an oversimplification. The author herself notes, immediately following the first quotation, that 'What this perceived prominence is due to, however, is a highly complex question which has caused a considerable amount of discussion in past years'. In other words, the phonetic causes, or correlates, of stress have been described in different ways. But there has also been disagreement about the *linguistic function(s)* of stress, about the *number of levels* necessary for an adequate description of stress in (for example) English, and about the *relevant domain* (the syllable, the word, larger domains).

Furthermore, what must be assumed to be (more or less) the same phenomenon has been treated under different headings such as *stress, accent* and *prominence*. Most scholars distinguish between some or all of these, but the distinctions are often not the same from one scholar to the next, and neither is the terminology: one person's stress may be another person's accent. It would be a daunting task to set out to disentangle the terminological and conceptual multifariousness which surrounds this topic today, and my goal here is somewhat more modest. In the following sections I will give a brief account of stress, accent and prominence in English, and where relevant also related languages such as German, Dutch and Swedish, as these have been presented in some of the works and schools of thought – both older and more recent – which are most relevant for my investigation. The special issue of terminology, as it will be employed in my own investigations later in this thesis, will be treated at the end of this chapter (in Section 1.8), after the general concepts have been outlined.

1.2 Earlier accounts of stress

Daniel Jones provides the following definition of stress in *An outline of English phonetics*: '*Stress* may be described as the degree of force with which a sound or syllable is uttered. It is essentially a subjective action' (Jones 1918: 245). That is, stress is defined with reference to the effort which the speaker makes to produce it, namely to increase the force of the articulatory action. The normal auditory result of this increased effort is *loudness*, it is stated, but this is not a necessary characteristic. Jones remarks in a footnote that stress may even fall on a silence, as in the abbreviated form [ḳkjuː] of *Thank you*. In such cases, 'A hearer familiar with the language would not perceive the stress objectively from the sound [...], but he perceives it in a subjective way; the sounds he hears call up to his mind (through the context) the manner of making them, and by means of immediate "inner speech" he knows where the stress is' (Jones 1918: 245, footnote 1).

This focus on the subjective action of the speaker and the hearer's attempt to reconstruct the speaker's effort makes Jones' definition quite different from the description of stress as relative prominence in the quotation which opened this chapter. In fact, although Jones states that stress will under normal circumstances (unlike the [ḳkjuː] example above) give added prominence to a sound (segment) or syllable, stress and prominence are not directly linked in his description. In *The Pronunciation of English* he states that 'Stress is not the same as "prominence" [...]; stress is one of the factors that may cause or help to cause a sound or syllable to be "prominent"' (Jones 1909: 141). The other factors which can make a sound more prominent are 'inherent sonority, length and intonation' (Jones 1909: 142). The distinction we find here between stress, sonority, length and intonation as contributing factors to the perception of prominence is similar to the distinction between loudness/intensity, sound quality, duration and pitch which is found in many modern accounts of stress and accent, but there are some crucial differences. To Jones, sonority, length and intonation are not part of the system of stress and their function is quite separate from that of stress.

The function of stress is to signal the important parts of an utterance: '[...] the relative stress of the words in a sequence depends on their relative importance. The more important a word is, the stronger is its stress' (Jones 1918: 262). That is, in addition to the lexical function of stress whereby the location of the stress in a word can be used to differentiate word meaning, stress has the function of marking some words as particularly important in a sentence. These are (normally) 'nouns, adjectives, demonstrative and interrogative pronouns, principal verbs, and adverbs' (Jones 1918: 262), that is, those normally referred to as lexical words or content words, while grammatical words, or function words, are unstressed.

Prominence, on the other hand, is used about the variation which characterises the difference between syllabic sounds, that is, the 'peaks' of prominence, and non-syllabic sounds, the 'valleys' of prominence. A quoted example is that 'In *button-hook* there are three peaks of prominence and therefore three syllables, the syllabic sounds being ʌ, n and u' (Jones 1918: 55). This is not the only use of the word, however. Later

in the same book Jones writes: 'When it is desired to give emphasis to a particular word in a sentence, that word has to be said with greater *prominence* than usual' (Jones 1918: 297, my emphasis). The choice of the word prominence rather than stress indicates that emphasis is not a part of the system of stress, that is, emphasis is not to be regarded simply as (very) strong stress, but as strong stress plus increased length and/or special intonation (meaning pitch modification). In fact, it is stated that of these '[...] intonation is the most important; it is generally, though not necessarily combined with extra strong stress on the emphatic word' (p. 297). It is interesting that Jones distinguishes the function of '... giv[ing] emphasis to a particular word in a sentence', which is done by means of increased *prominence*, from the function of marking the important words in a sentence, which is done by means of stress. It should be noted, though, that Jones is not consistent about this distinction and sometimes he uses the term *stress* for the means with which we signal emphasis: 'When it is desired to *emphasise* a word for contrast, its stress is increased, while the stress of the surrounding words may be diminished' (Jones 1918: 264). It would appear that Jones does in fact not mean stress, as he has defined it, in this case but prominence, especially considering his statement later in the same book that 'Contrast-emphasis is expressed mainly by intonation' (Jones 1918: 298). There are several reasons why the function of marking some (namely lexical) words in a sentence as 'important' might be seen as separate from giving emphasis to a particular word. While the first function seems to be a necessary part of every sentence, or utterance, even in a neutral context, emphasis can be seen as something 'extra' which is added under certain circumstances. Furthermore, some phonetic features, especially variations in pitch (Jones' *intonation*), are always involved in marking emphasis, while they are not always involved in marking 'important words'. Uniting these variations in both form and function in a coherent frame can be done in different ways (as it will appear from the following sections), and it is a testament to the complex nature of these phenomena (stress, prominence, emphasis) that it is difficult to maintain complete terminological consistency in the description of their functions on different levels or across different domains such as the segment, syllable, word and sentence.

With regard to *degrees of stress* Jones argues in *The phoneme* (1950) that only two degrees (stressed or unstressed) should be distinguished at the word level, that is, the lexical function of stress, and that '[...] the so-called intermediate degrees of "stress" are as a rule degrees of *prominence* due to tamber, sound-groupings, length or voice-pitch, or combinations of these, with or without the accompaniment of stress' (Jones 1950: 148), although elsewhere he operates with secondary stress in longer (non-compounded) words. Sentence level stress, however, seems to be considered completely scalar, and he refers to various degrees such as 'strong stress', 'really strong stress', and 'medium or fairly strong stress' (Jones 1918: 299), suggesting that there are no fixed levels.

The definition of stress as breath force is also found in most other accounts from that period (the first half of the 20th century) such as Jespersen (1899: 352), and Pike (1943: 119), who uses the phrase 'stronger initiator pressure' to characterise

stressed syllables. Later Abercrombie (1967: 35) states that 'A syllable produced by a reinforced chest-pulse is called a *stressed* syllable'. There seems to be common agreement on this position on stress until the 1950s when it was challenged by especially Bolinger and Fry, but also by Gimson in his article 'The linguistic relevance of stress in English' (1956). Here Gimson argues that the description of (linguistic) stress as signalled mainly by breath force is partly due to a historical shift in terminology from the term 'accent', which 'included hitherto a variation of pitch or intensity as a means of rendering a syllable prominent' (Gimson 1956: 144) to 'stress', causing a shift in focus from pitch to intensity, or loudness. However, the linguistically relevant feature to Gimson is the relative prominence of the syllables, irrespective of the phonetic means by which it is achieved, and he cites early experimental work as showing that 'The inefficiency of stress [i.e. "extra expiratory effort or extra loudness"] as a sole means of achieving prominence has been well illustrated by N. C. Scott' (Gimson 1956: 147). Scott (1939) showed that when pitch cues are neutralised English listeners (11 students) found it difficult to distinguish the verb and noun forms of the word *imports*.

Gimson also criticises the view that stress is essentially a subjective action on the part of the speaker and states that '[...] only those sound features are worthy of consideration which are capable of being perceived by a listener' (Gimson 1956: 143). He does not deny the speaker reality of a stress on the first [k] in Jones' example of *Thank you*, realised as [ˌkkjuː], but does deny the linguistic relevance of such subjectively felt stress. Similar criticism was raised a few years earlier in Jassem (1952) and repeated in Jassem and Gibbon (1980), stating that 'whatever cannot be heard by a normal human ear *ipso facto* lies outside the field which is covered by phonetics as a strictly linguistic discipline' (Jassem 1952 in Jassem and Gibbon 1980: 4).

The significance of this shift in perspective from subjectively felt speaker action to hearer perception is that the defining elements of stress become the features which are most perceptually salient, that is, result in the greatest perceived prominence, rather than the features which may be most easily felt by the speaker when producing stress – which presumably is breath force. It had already been suggested in the above-mentioned papers by Jassem and Gimson and by many other writers, including Jones, that tone, or pitch, could be a very efficient way to signal prominence (in addition to duration and sound quality), and since the 1950s much experimental work has been carried out to determine which factors are mostly responsible for creating a sensation of prominence. Some of this work will be described in the next section.

1.3 Early experimental work on stress

The above-mentioned investigation by Scott (1939) indicated that intensity might not be as efficient a cue to the perception of stress as expressed by many phoneticians, and in the 1950s and early 1960s Fry carried out a series of experiments which examined the relative contributions of F_0, duration, intensity and formant structure, or spectral composition, to the perception of stress. In all the experiments he

examined to what extent manipulation of these features could shift listeners' perception of words such as *subject, digest* and *permit* from verb to noun or vice versa, using synthetic speech stimuli. The synthetic stimuli were based on recordings of American English, and were presented to both British and American listeners in some of the experiments, with no marked difference between the groups (Fry 1958a: 134). In Fry (1955) he demonstrated that both intensity and duration can act as cues to stress, but that duration is a more effective cue than intensity. In Fry (1958b) and Fry (1958a) he compared F_0, intensity and duration, and again the results showed that duration is a more effective cue than intensity. However, changes in F_0 turned out to be an even more efficient cue. It differed from the other two cues in being a question of either-or rather than more-or-less. The presence of a pitch change was important; the size of the change was less important. Fry suggests that '... sentence intonation [corresponding to F_0 change within one syllable] is an over-riding factor in determining the perception of stress and [...] in this sense the fundamental frequency cue may outweigh the duration cue' (Fry 1958a: 151). It is important to realise that in all of Fry's experiments the test word under investigation (*subject, digest*, etc.) was placed at the end of the carrier sentence, e.g. 'Where is the accent in <test word>', or as single word utterances. The test words were therefore always in nuclear position, which may have a great influence on tonal possibilities, and listeners' expectations of these, for any particular word. The last experiment (Fry 1965) suggested that formant structure (F1 and F2 values) is in fact a less efficient cue than intensity, although he hesitated to draw too strong conclusions from the results. These experiments established on a solid experimental foundation a hierarchy of stress cues which has found widespread acceptance among phoneticians and is often mentioned in phonetics textbooks, although it has not gone unchallenged by later research (see e.g. reference to Nakatani and Aston 1978 and Berinstein 1979 later in this section). This hierarchy of perceptual and acoustic cues is as follows:

perceptual cues	pitch	> length	> loudness	> quality
acoustic cues	F_0	> duration	> intensity	> formant structure

Another writer who was working along the same lines at the time was Bolinger, who published a series of articles which set out to demonstrate that intensity is a poor cue to stress, and that other properties such as disjuncture (Bolinger and Gerstman 1957) and particularly pitch (Bolinger 1955, Bolinger 1958) are far more important cues. Bolinger based his criticism of the 'stress equals loudness' school of thought partly on the experiments of others, such as Fry (Bolinger and Gerstman 1957: 246), and partly on experiments carried out by himself and his co-workers. In the most well-known article (Bolinger 1958) he reports the results of a series of tests which all point to the primacy of pitch over intensity or duration as a cue to perceived promi-

nence, and proposes to change not only the definition of stress but the term itself:

> Having given up the more usual definition of stress, I think it is wise, because of association, to give up the term also. From this point on I shall therefore refer not to stress but to PITCH ACCENT, or simply ACCENT, meaning prominence due to the configuration of pitches (Bolinger 1958: 127).

To Bolinger pitch obtrusions are responsible for marking the prosodically and semantically meaningful events, referred to as 'semantic peaks' in for example Bolinger (1961: 84.) The role of duration and intensity is not completely denied, but they are regarded as secondary or ancillary cues, which may help (particularly duration) to disambiguate which of two possible pitch accented syllables actually carry the accent (Bolinger 1958: 138-39).

Bolinger mentions one other factor which might be involved in the perception of prominence, or stress, namely *position*:

> It is conceivable that stress is climactic, and that we attribute extra intensity to the position at the end, even when it lacks it phonetically. [...] If position overrides pitch, which in turn overrides intensity, we have one explanation of why the end stresses are so consistently marked as 'louder' (Bolinger 1958: 125).

Bolinger then quotes results from an experiment ('Test 8') which suggest that the latter of two pitch prominent syllables will be heard as the most prominent one, unless both pitch and duration cues overwhelmingly point to the first syllable. The listener task in the experiment was to identify a synthetic speech string as either '... the word *úndertaking*, "what a mortician does," or *undertáking*, "enterprise"', that is, to locate the main stress in the compound. When both the potentially stressed syllables (*un-* and *-ta-*) were marked by a pitch rise the majority of listeners perceived the syllable *-ta-* as carrying the main stress, even when the pitch excursion on *un-* was far greater than on *-ta-*. Bolinger's reference to *intensity* and *loudness* in the quotation above is somewhat odd, however, since the crucial matter seems to be the presence, or not, of a second prominent syllable to compete with the first. The results may be taken to indicate that the last (pitch) prominent syllable will be understood as the main stress unless very strong phonetic cues point to an earlier position. This interpretation would be very much in line with the findings of Brown *et al.* (1980), that the final prominent syllable in an utterance (intonation unit) tends to be heard as the nucleus (their findings will be treated in more detail later).

The issue of the relation between stress cues and position in the utterance or phrase has not in general played a large role in the literature on stress, although a few studies, in addition to those mentioned above, have indicated either directly or indirectly the importance of this factor. Adams and Munro (1978) investigated production and perception aspects of the correlates of stress for both native speakers of (Australian) English and a group of non-native speakers (native speakers of various Asian, purportedly syllable-timed, languages). They specifically wanted to find

correlates of stress in connected speech, as opposed to the single word utterances or nuclear position only which had so far been predominant in research on stress. They did not use spontaneous speech but a series of read texts including nursery rhymes and verse with very clear rhythmic patterns, but also more 'normal', that is, rhythmically less regular, prose. They found that the most frequently used cue to stress was duration, ahead of pitch and with amplitude (intensity) as the least frequently used. The significance of this finding is that when stress is examined in all positions in an utterance, pitch no longer stands out as the predominant cue, in terms of frequency of use. This does not necessarily indicate, of course, that it is not as efficient a cue as had been established by Fry, Bolinger and others – only that it is perhaps not used to the same extent in all positions of an utterance. Incidentally, Adams and Munro (1978) also found that while the native and non-native speakers differed with regard to the *placement* of stress, there were no consistent differences in how they *signalled* stress, that is, which acoustic parameters were used. Nakatani and Aston examined how the perception of (actualised lexical) stress was affected by the acoustic parameters duration, pitch, amplitude and vowel quality, but also looked at the influence of position in the utterance (Nakatani and Aston 1978). They used a combination of reiterant speech, replacing the word under investigation with the syllables 'mama' in each utterance, and a type of speech synthesis which allowed them to vary, or manipulate, the acoustic parameters which were hypothesised to be responsible for cueing stress. Using this method provided a large degree of control over the features under investigation. Their results showed that the perceived stress pattern of a 'mama' word was influenced by pitch, duration and vowel quality, but not by amplitude. The effect of the first three features was additive, that is, the sum of the features accounted well for perceived stress with relatively little interaction between the cues. Duration was found to be the best cue overall, pitch the second best cue and then vowel quality. This is in good agreement with the findings of Adams and Munro (1978) and another challenge to the claim that pitch is the most important cue to stress. They also found some interesting interactions between the acoustic cues and what they refer to as the linguistic factors, namely position in the utterance and accentuation, defined by them as emphasis or contrastive stress (Nakatani and Aston 1978: 1). They mention two of these in their summary of findings, namely that 'duration was nullified as a stress cue for sentence final words, and pitch was nullified as a stress cue for words after an accented [i.e. emphatic] word' (Nakatani and Aston 1978: abstract). In other words, in nuclear position lexical stress was not signalled by variation in duration, which was presumably neutralised by phrase final lengthening, and in general no pitch excursions were found in post-emphatic position. But there were other important connections between the acoustic cues and position, which may have been obscured somewhat by the way in which they defined position. Their definitions relied in part on grammatical phrase structure, so their 'phrase initial' context was 'verb in verb phrase' as in

The mayor <u>lectured</u> around the city

'Phrase medial' context was 'adjective in noun phrase':

The crippled photographer took the picture

And 'phrase final' position was 'noun in noun phrase'

Our favorite actress was on television

One of their contexts was based on sentence structure rather than phrase structure, so that 'sentence final' means 'final noun in sentence'

My father plays the trumpet.

All examples are from Nakatani and Aston (1978: 32). The underlined words are the ones which were replaced by 'mama' in the actual test. Nakatani and Aston found a strong connection between the shape, or distinctness, of pitch contours and their efficacy as stress cues and noted that 'differences in the pitch contours of 'MAma' and 'maMA' words were most evident for words in phrase-medial and sentence final contexts; pitch was the best stress cue for words in these contexts' (Nakatani and Aston 1978: 19). This is also apparent from their Fig. 4. In other positions pitch was superseded as a cue by duration and sometimes even vowel quality. While it may be difficult to generalise on the basis of Nakatani and Aston's definitions of position, it becomes much easier if position is regarded as a function of prosodic structure: the words in phrase medial and sentence final position are the first and last lexical items of the utterance, respectively, so the obvious generalisation is that pitch is the most efficient stress cue near utterance boundaries, or perhaps more appropriately, near major phrase boundaries. One important side effect of such a definition or generalisation is that the presence or absence of prosodic boundaries internally in the utterance becomes highly significant. For example, a phrase boundary might be expected (optionally) before the adverbial in the 'phrase initial' context, or between subject and predicate in the 'phrase final' context. However, whether such utterance-internal phrase boundaries would have a large effect on the pitch contour and thereby on the importance of pitch as a stress cue cannot be determined from the results from Nakatani and Aston (1978).

Huss (1978) looked at lexical stress patterns in minimal stress pairs such as *ím*port (n.) – im*pórt* (vb.) and similarly *insult* and *decrease*, in post-nuclear position. In all his sentences the nucleus was used for explicit contrastive emphasis, so post-nuclear also means post-contrastive in this work. He found that lexical stress distinctions were neutralised in post-nuclear position. There were no differences in pitch between the word pairs in this position, and listeners tended to perceive the stress pattern according to a pre-established rhythm in the utterance: they would perceive the test word as either iambic or trochaic in accordance with the rhythmic pattern of the preceding part of the utterance. Huss did find systematic and statistically significant differences in both intensity and duration between the word pairs, but these acoustic cues were overruled by rhythm in his test. This finding is contradicted in Nakatani and Aston (1978), where listeners did succeed in identifying the correct rhythmical pattern in post-nuclear position, even in the absence of pitch cues. Both Nakatani and Aston's results and Huss' own production results indicate that the distinction is at least encoded in post-nuclear position, even if it is not always perceived

by listeners.

While most of the studies of the perceptual and acoustic correlates of stress have either assumed that these are universal or have only been concerned with the particular language variety under investigation, one study set out to demonstrate that stress cues are bound by phonological constraints in a language (Berinstein 1979). More specifically, if a language uses an acoustic dimension, such as F_0 or duration, for phonemic distinctions, this parameter will be less important as a stress cue in that language. If a language has phonemic tone distinctions, pitch will not be used (to any large degree) as a stress cue, and if a language has phonemic length, duration will not be a good stress cue. Comparing (American) English and the two Mayan languages K'ekchi and Cakchiquel, Berinstein demonstrated that while duration was a good stress cue in English, which does not have phonemic length (at least within a certain phonological analysis, as adopted by Berinstein), it was not a cue to stress in K'ekchi which does have phonemic length. Cakchiquel is similar to K'ekchi in having fixed final stress, but does not have phonemic length and the fact that duration *was* used as a stress cue in this language was taken as confirmation that the inefficiency of duration as a stress cue in K'ekchi is linked with phonemic length distinctions in this language (Berinstein 1979: 44). It is somewhat uncertain how this would affect our expectation of the use of stress cues in British English versus Danish. First of all, it is debatable whether British English has phonemic length distinctions. While all consonants are regarded as short, we normally recognise phonemic vowel length distinctions, although these are also accompanied by clear differences in vowel quality (Gimson 1989). Danish definitely has phonemic vowel length, as demonstrated by such pairs as 'hvile, ville' ['viːlə, 'vilə] (*rest, would*) and 'læse, læsse' ['lɛːsə, 'lɛsə] (*read, load*). In some contexts stress reduction leads to the complete loss of phonemic vowel length including the propensity for *stød*, but in other contexts long vowels may retain stød and only be partially shortened, that is, not identical to the corresponding short vowel (Fischer-Jørgensen 1984). This seems to speak against Berinstein's conclusions about duration as a stress cue in languages with phonemic (vowel) length, but it should be mentioned that only production data are available for Danish. The connection between the acoustic cues and perceived stress in Danish is largely unexplored, except for a study by Thorsen, which showed that relatively minor differences in the timing of a rise in F_0 can, in certain contexts, be sufficient to shift the perception of the location of stress between first and second syllable in a minimal stress pair such as 'billigst – bilist', ['bilisd – bi'lisd] (*cheapest – motorist*) (Thorsen 1982).

1.4 More recent investigations

Although stress, or prominence, has not been given the same attention as intonation in recent years, there have been a number of studies on prominence in not only English, but also other languages such as German, Dutch and Swedish. Many of these studies have been connected with the requirements of speech technology applications: either the ability to automatically find stressed syllables in speech

recognition systems (Wightman and Ostendorf 1994, Streefkerk 1997, Streefkerk, Pols and ten Bosch 1999) or finding specifications for coding stress in speech synthesis systems (Heuft and Portele 1996, Wagner 1999), often using large corpora as the basis for analysis. However, there have also been some more 'traditional' studies of prominence not only directed towards speech technology applications, and often using more controlled types of speech (Heldner 2001a, Sluijter 1995, Terken 1991, Gussenhoven, Repp *et al.* 1997). Whichever method is used, these studies, like previous ones, typically attempt to find the most important or relevant acoustic features to account for the perception of stress or prominence.

1.4.1 Pitch and duration

Naturally, variation in F_0 receives a fair amount of attention, partly because of the findings of Fry and Bolinger, as reported above, but also because the prosodic models which are used for the studies are often rooted in the Autosegmental-Metrical (AM) tradition, such as the ToBI framework (see Section 1.6) in which the 'pitch accent', defined by reference to pitch excursion, is considered the relevant unit in an account of the location of prominent syllables, if not the actual degree of prominence of these. This is partly reflected in the assumption made in Wightman and Ostendorf (1994: 472) that 'there is always some type of pitch accent associated with a prominent syllable', although their analysis also focuses on other acoustic cues such as duration and intensity.

Among the studies which pointed to the significance of F_0 in the perception of prominence is A. C. M. Rietveld and Gussenhoven (1985), who demonstrated that a difference in pitch excursion size of as little as 1.5 semitones could create a difference in perceived prominence if the two pitch peaks were within the same pitch range. The account of prominence in Liberman and Pierrehumbert (1984) also points to a direct relation between pitch height and prominence levels. Their Figure 4 shows pitch contours of seven productions of the phrase *Anna*, with variations in overall emphasis that are clearly reflected in the peak height of the pitch contour (Liberman and Pierrehumbert 1984: 159). Prominence differences are described as quantitative, 'That is, the underlying parameter is continuously variable' (Liberman and Pierrehumbert 1984: 161). In other words, the prominence level of a pitch accent is reflected in its peak height relative to other accents. However, as they also acknowledge, this relation is not a simple one but depends on the position of the accent in the phrase, because of the differences in F_0 range found at the beginning of a phrase (wider range) and the end of a phrase (narrower range) – the phenomenon often referred to as *declination*. It had been demonstrated by Pierrehumbert (1979) that listeners compensate for the expected tendency for F_0 to fall gradually through the course of an utterance and perceive an F_0 peak of a given value as higher in pitch when it occurs later in the utterance. Or stated differently, for two F_0 peaks to be perceived as having the same pitch height the second must be lower than the first. These results are referred to in both Liberman and Pierrehumbert (1984) and A. C. M. Rietveld and Gussenhoven (1985) as reflecting an effect on the perception of

prominence, although the issue addressed in Pierrehumbert (1979) was pitch height and not prominence. This may be taken as an indication of the assumed direct relation between pitch height and prominence. This assumption was tested (for Dutch) in Gussenhoven and Rietveld (1988), where raters were asked (in one of the experiments) to judge the prominence of the second of two F_0 peaks with varying peak heights. The results showed that listeners do take declination into account when judging prominence: the second of two peaks with the same F_0 was judged to be more prominent than the first one. Again, this result is taken as confirmation of the findings of Pierrehumbert (1979), although it might be better to regard it as an extension to those findings. Terken (1991) examined the connection between declination and both pitch height and prominence in separate but parallel experiments and found that when listeners are asked to match two pitch accents in an utterance the required difference is larger when judging prominence than when judging pitch height. He concludes that listeners use different strategies in judging prominence and pitch (Terken 1991: 1773); the relation between the two is perhaps not as simple and straightforward as suggested by some of the previous research: a function of the F_0 maxima in the utterance in relation to the baseline declination. Instead he suggests that the connection is complex and that an explanation needs to incorporate both local and global characteristics of the intonation phrase.

The relation between prominence, F_0 peaks and declination were further examined in Gussenhoven, Repp *et al.* (1997) in five experiments which tested both local and global parameters. Among the results were the finding that an F_0 peak of a given value is perceived as more prominent later in an utterance, regardless of whether there is an earlier F_0 peak to compare it with, which was the context used by Pierrehumbert (1979). They also found that the height of the F_0 value at the end of a falling contour does not affect the perceived prominence of the peaks in the contour (so final F_0 values, or F_0 'offset' is not used as an anchor point), but that the F_0 height of initial unstressed syllables in an utterance can act as a reference, or anchor point, with regard to pitch and prominence and affect the evaluations, but only if the duration of this 'onset' exceeds 400 ms.

Pitch, or F_0, has also featured prominently in other investigations, such as the model for synthesising German prosody described in Heuft and Portele (1996), where duration is also included, or an early paper by Streefkerk and associates (Streefkerk and Pols 1996) which describes the relation between perceived prominence and pitch movements in read Dutch speech and states that 85.9% of all the prominent accents in the investigations were associated with pitch movement. However, in later papers other acoustic parameters are also included, such as duration and intensity (Streefkerk 1997, Streefkerk, Pols and ten Bosch 1998, Streefkerk, Pols and ten Bosch 1999). While it was found in Streefkerk, Pols and ten Bosch (1998) that perceived prominence correlated well with F_0 variation and intensity they found a fairly poor correlation between duration and prominence. This is in contrast to most other investigations of the acoustic correlates of stress, or prominence, such as Nakatani and Aston (1978) or Silipo and Greenberg (1999, see below). In Streefkerk,

Pols and ten Bosch (1997) it was found that listeners were able to identify prominent words even in the absence of F_0 cues, although much less confidently so. The number of items with near total agreement among raters fell drastically when pitch was excluded as a cue, and there were (other) indications that listener behaviour differed with regard to the reliance on pitch. It was also found, perhaps not surprisingly, that listeners found it easier to mark prominence on words rather than on syllables, and that the results from the word prominence task represented sentence level prominence, or accents, better than the results from the syllable prominence task.

Some investigations have found that F_0 is not always a very efficient or important cue to stress. In Silipo and Greenberg (1999) two transcribers (both trained linguists) assigned stress (primary, intermediate and no stress) to a corpus of spontaneous (American) English dialogue. Their assignments were then compared to the results of an automatic stress detection algorithm using the parameters duration, pitch and amplitude and combinations of these. The degree of agreement between the scores of the automatic algorithm for various parameters and the human transcribers can be taken as an indication of how important those parameters were for the transcribers. The results showed that one of the transcribers relied primarily on duration while the other transcriber used both amplitude and duration as (equally important) cues in the assignment of primary stress. The pitch parameter gave the worst performance of the automatic algorithm compared with the transcribers. It should be pointed out, though, that a very simple measure of pitch was used in the algorithm, namely the average value of F_0 within the vocalic nucleus. This problem was addressed in Silipo and Greenberg (2000) in which the experiment was repeated using F_0 range in addition to the parameters from the previous study, as well as certain measures derived from all four parameters. Again, duration proved to be the most important parameter, and the product of duration and amplitude was the most important combination. F_0 range did turn out to be more efficient than average F_0, but it is argued in the article that the variation in F_0 range is strongly connected with, and can be explained away with reference to, variation in duration. Hitchcock and Greenberg (2001) attempt to show that stress accent in the same corpus is associated with vowel height, so that open vowels are more likely to be perceived as stressed than close vowels, partly through the connection between vowel height and vowel duration. However, the article does not contain information about a possible connection between vowel height and lexical incidence, for example, the possibility that many (half-)close vowels are found in very frequent grammatical words, which are typically unstressed. This might explain a large part of the observed connection between vowel height and perceived stress.

Pitch has often been described as particularly important in marking higher levels or degrees of prominence, such as emphasis or focus (Jones 1918, Kingdon 1958a), but some studies have shown that pitch excursions are not always necessary even under these conditions. It has been found for Swedish that an F_0 excursion, the focal accent rise, is a reliable correlate of focus and is strongly associated with focused words: it is present when a word is focused and not present when a word is

not focused (Heldner 1996). However, it is reported in the same paper that listeners are able to detect focused words even when F_0 cues have been removed and conversely that adding F_0 excursions to non-focused words is (generally) not sufficient to make these words be perceived as focused. Heldner therefore concludes that F_0 is neither a necessary nor sufficient cue to focus in Swedish. Other acoustic parameters which were investigated included segment duration, overall intensity and spectral tilt, and of these only duration was found to correlate with focus in all (the tested) positions in the utterance. Although the importance of F_0 movements is not denied in Heldner (1996), it is shown that F_0 is not such a strong cue that it can override other, conflicting, cues, most importantly duration, and that an account of focus therefore needs to take these into account as well. (At the end of the paper Heldner speculates that an account of focus should be based not only on local features but also on global features such as pitch range and phrasing.) The conclusions from Heldner (1996) are challenged somewhat in Sautermeister and Eklund (1997) where the relative contribution of F_0 and duration to the perception of prominence in Swedish are compared. Using reiterant resynthesised speech it is shown that F_0 is used as the primary cue to prominence when both F_0 cues and duration cues are present and competing. The listeners were asked to indicate which of a series of syllables was perceived as the most prominent one. They did find that in the absence of F_0 cues there was a general tendency to perceive lengthened syllables as more prominent, but they found that not all listeners exhibited this behaviour. So in this study, and at least for some listeners, F_0 does seem to be a necessary cue to the perception of prominence.

1.4.2 Spectral tilt/balance/emphasis

One other acoustic parameter (in addition to F_0, duration and intensity) which has received much attention since the mid 90s, especially through the work of Sluijter and van Heuven, is spectral tilt, also referred to as spectral balance or emphasis (Sluijter 1995, Sluijter and van Heuven 1996, Sluijter, van Heuven and Pacilly 1997, Sluijter, Shattuck-Hufnagel *et al.* 1995). Spectral tilt is the distribution of energy in different frequency bands, and an increase in energy in the higher frequency bands is seen as a correlate of stress. The underlying explanation is that stress is associated with greater vocal effort, which results in a different glottal pulse with a sharper trailing flank, again resulting in an increase in energy in the higher frequency bands but not (or hardly) affecting the lower frequency bands (Sluijter 1995: 6). In other words, the physical effort with which a syllable is uttered is reflected in the steepness of the slope which characterises the decrease of energy with an increase in frequency. This association, if sustained, can be seen as a restoration of the claim made by e.g. Jones, Jespersen and Pike (see above) that stress can be described as 'the degree of force with which a sound or syllable is uttered' (Jones 1918: 245). Sluijter and van Heuven use the presence and effectiveness of this acoustic cue to promote the view that pitch accent and stress are two formally and functionally separate categories. Pitch accents, that is, syllables with an accent lending pitch movement, in their view

indicates focus (of smaller or larger scope); linguistic stress, that is, the location of the lexically stressed syllable, is indicated by means other than pitch movements when the word is non-focused (Sluijter 1995). In Sluijter and van Heuven (1993) and Sluijter and van Heuven (1996) they show that spectral tilt, that is, the distribution of energy in higher frequency bands (0.5–1, 1–2, 2–4 kHz) is a strong acoustic correlate of stress in the target words 'canon' /ˈkaːnɔn/ (*cannon*) – 'kanon' /kaːˈnɔn/ (*canon*), as well as reiterant versions /naːnaː/ of these, in the carrier sentence 'Wil je [target] zeggen' /ʋɪl jə [target] zɛχə(n)/ (*Will you [target] say*) either with pitch movement on the stressed syllable of the target word or on the stressed syllable of the following word. They concluded that spectral tilt is 'a clear acoustic correlate of stress [...,] even more reliable than overall intensity [... and] close in strength to duration' (Sluijter and van Heuven 1996: 2481). In Sluijter, van Heuven and Pacilly (1997) they demonstrate that spectral tilt is also a very efficient *perceptual* cue to stress. They used resynthesised reiterant versions of the phrase from the production study mentioned above, and varied duration, overall intensity and intensity in the higher frequency bands (above 500 Hz) in seven steps. In line with previous research they found that overall intensity was a poor cue to stress, while duration was a good cue. Spectral tilt also proved to be a good cue, and in the test condition where stimuli were presented through a loudspeaker it was in fact more efficient than duration, due to a decrease in the effect of duration as a cue under these conditions. It was hypothesised, and confirmed in a subsequent experiment, that reverberant conditions, such as might be present in a normal room, may disguise temporal information related to segment boundaries (Sluijter, van Heuven and Pacilly 1997: 508). The fact that such reverberant conditions are found in many normal speech situations is used as support to the stated importance of spectral tilt as a cue to stress.

Campbell and Beckman (1997) investigated spectral tilt in American English and found an increase in energy in the higher frequency bands when the syllable carried nuclear accent, but they found no systematic differences between stressed and unstressed syllables in the absence of an accompanying pitch accent. They use these findings to support their claim that pitch accents marking focal prominence are phonated in a special, more emphatic way, and to reject the idea that linguistic stress is a separate, independent level marked systematically by spectral tilt, at least in (American) English.

The connection between spectral tilt and focal accent has also been investigated for Swedish, where it is usually referred to as spectral emphasis (Heldner 2001b, Heldner 2003, Heldner 2001c). Spectral emphasis is shown to be a good acoustic correlate of focal accent; in many cases it is better than overall intensity, especially because it is more robust to variations in the position of the word in the phrase and segmental influences (Heldner 2003, Heldner 2001c). However, an attempt to demonstrate that spectral emphasis is also a good perceptual cue to prominence, and thereby accent, was unsuccessful (Heldner 2001b). Focal accents in which the spectral balance had been manipulated (by increasing the amplitude of frequencies above the fundamental and decreasing the amplitude of the fundamental) were not

consistently perceived as more prominent by listeners. And furthermore, introducing spectral emphasis as an accent/stress marking feature in synthetic speech did not improve the perceived quality of synthesised utterances, although the production studies had suggested a strong correlation between spectral emphasis and prominence. Several possible explanations of these negative results are suggested in Heldner (2001b), one being the fact that the manipulations were carried out on words which were already accented and therefore already had the required spectral emphasis. The positive results obtained in Sluijter, van Heuven and Pacilly (1997) concerned stressed but not accented words, that is, what one might call a lower degree of accentuation. The conclusion must be that while spectral tilt (balance, emphasis) has been demonstrated to be a strong acoustic correlate of stress and focal accent, its relevance as a perceptual cue remains somewhat uncertain.

1.4.3 Prominence and word class

Some studies have shown a connection between prominence and word class, or part of speech. It is of course well known that lexical words tend to be stressed and grammatical words tend to be unstressed, as was also clear in the quotation from Jones (1918) in the beginning of this chapter, but some systematic variation within these two categories has also been observed. In Streefkerk, Pols and ten Bosch (2001) ten naive listeners were asked to indicate the prominent words in a corpus of 1244 Dutch sentences. Prominence markings were binary, that is, prominent or non-prominent. The cumulative scores of the ten listeners were then used as indications of measured prominence level. Words were classified in 11 categories and the mean prominence level for each category was calculated. The results show a clear distinction between grammatical and lexical words with no overlap. Negations and adjectives were the most prominent lexical words, with ratings of approximately 6 on the scale from 0 to 10, followed closely by numerals and nouns. Adverbs were considerably less prominent (3.8), but the least prominent lexical words were verbs (2.8). Contextual factors, such as the position of the words in the clause or phrase, are only briefly touched upon, but it is noted that the four most prominent categories are generally more prominent in initial position than elsewhere in the sentence, and nouns are considerably less prominent when they are preceded by an adjective. These observations were used in the subsequent attempt to assign prominence automatically (using Feed Forward Nets), but the article does not deal with the linguistic significance of the differences in prominence levels.

Similar results, but for German, are presented in Widera *et al.* (1997). Here three listeners assigned prominence to every syllable of 6434 words on a scale from 0 to 31. Again, lexical words were found to be more prominent than grammatical words, and verbs were, also in this experiment, perceived to be the least prominent lexical words. The connection between prominence and the position of the word in the clause is investigated more systematically in this study, which distinguishes five positions: first, second, third, medial and last. The results show a correlation between the prominence value of a particular word class and position, and it is noted that

'prominence values tend to be increased by about 4 points [on the scale from 0 to 31] in clause initial and clause final positions' (Widera *et al.* 1997: 999).

Both Streefkerk, Pols and ten Bosch (2001) and Widera *et al.* (1997) point to a connection between perceived prominence and word class and position in the clause. Unfortunately, very few details about the perceived prominence levels and position are reported, and neither paper discusses the very obvious possibility of a connection between word class and position. Depending on the material that has been used, especially the semantic and syntactic structure of the sentences, there may be extensive inter-dependence between the two; for example, many simple declarative sentences have a subject with one lexical word followed by the verb and then (the rest of) the predicate, as in the English sentence

Peter asked for the bill

An observed variation in perceived prominence between the proper noun *Peter* and the following verb *asked* could be caused by the difference in word class, but also by the difference in position in the clause or phrase. Furthermore, since it may in fact be position in the *phrase* rather than in the *clause* which is the relevant parameter, one must also take phrasing in longer utterances into consideration. And here too there may be a connection between position and word class, since there is a tendency for prosodic boundaries to coincide with boundaries between higher-level syntactic constituents such as clauses (Hirst and Di Cristo 1998: 36). A proper account of the connection between prominence and word class therefore needs to control for these factors, either through a detailed analysis of the prosodic context of each word, or through controlled experiments where other factors are kept constant or varied systematically.

1.4.4 Prominence scales

As it appears from the discussion of some of the above papers, including the two in the previous section, the issue of prominence levels has received very different treatment in different studies. Some are only concerned with identifying stressed as opposed to unstressed words or syllables and thus typically only use a binary scale (Wightman and Ostendorf 1994, Buhmann *et al.* 2002, Streefkerk and Pols 1996), or alternatively a three-point scale indicating no stress, primary stress and intermediate stress (Silipo and Greenberg 2000, Silipo and Greenberg 1999). However, in studies which are concerned with variations in perceived prominence, or stress, levels several strategies have been employed. Fant and Kruckenberg used a 31-point scale from 0 to 30 (Fant and Kruckenberg 1989, Fant, Kruckenberg *et al.* 2001). Listeners were asked to indicate stress graphically with pencil marks on a vertical line above the text, and were advised that typical values for stressed and unstressed were 20 and 10, respectively. It appears from their Fig. 8 in Fant and Kruckenberg (1989) that most, if not all, stress markings in the depicted utterance are within this range. Despite the large number of categories they claim a high degree of consistency from the 15 listeners, with standard deviations of around 2–4 scale points and an estimated

reliability of 0.5 to 1 scale point (Fant and Kruckenberg 1989: 14-15). The same scale has been used in several other studies, although sometimes listed as a scale from 0 to 31 but still credited to Fant and Kruckenberg (Heuft and Portele 1996, Widera *et al.* 1997, Wagner and Portele 1999). A similar method was used by Eriksson, Thunberg and Traunmüller (2001) Eriksson, Grabe and Traunmüller (2002), and Wagner and Fischenbeck (2002), where listeners could indicate prominence by adjusting sliders on a visual display on a computer screen to reflect the perceived prominence. However, some research has shown that while listeners tend to agree on the location of stressed words, agreement is much smaller with regard to the exact stress level to be assigned to a given word or syllable, especially when the listeners have not received extensive training in the labelling procedure used in the stress assignment task (Wightman 1993, Wightman 2002). An alternative method, proposed for example in Wightman (1993), is to use a simple binary labelling system and let the prominence level of each word or syllable be reflected in the number of listeners who assigned stress to this word. The advantage of this method is that the task becomes much simpler and can be accomplished by so-called naive listeners, that is, listeners who have not been trained in a particular tradition.

This last method has in principle been used in the work by Streefkerk and associates (Streefkerk and Pols 1996, Streefkerk, Pols and ten Bosch 1997, Streefkerk, Pols and ten Bosch 1998), but it is only in Streefkerk, Pols and ten Bosch (2001) that the resulting information is used for more than simply deciding on a threshold between accented and unaccented words. Whether one decides to use a simple binary system, a scale with a large number of levels or something in between it would appear that using the cumulative, or mean, scores of multiple listeners can remove or level out some of the disagreement between individual listeners and function as a good expression of inter-listener perception.

1.5 Stress and the British school of intonation

The intonation of Southern British English has been subjected to detailed descriptions throughout the twentieth century. Most of the descriptions have, in line with general English phonetic traditions, a clear pedagogical aim: they set out to explain the use of intonation in a manner which will enable the foreign learner to acquire the correct patterns of intonation. Most of this work has been based on introspection and examples of the use of intonation patterns collected by the authors themselves or previous descriptions (Palmer 1922, Armstrong and Ward 1931, O'Connor and Arnold 1973) while others are based on systematic impressionistic analyses of a large corpus of utterances (Crystal 1969). Although certain differences exist between the various descriptions, and some concepts have developed or changed over time, many share at least the basic concepts, and are often referred to as belonging to the *British school of intonation analysis* (Palmer 1922, Armstrong and Ward 1931, Kingdon 1958a, O'Connor and Arnold 1961, O'Connor and Arnold 1973, Crystal 1969, Gimson 1989, Cruttenden 1997).

The British school is mostly concerned with describing how certain intonational patterns, or intonation contours, relate to syntactic structures or semantic or pragmatic functions, that is, which patterns are appropriate for which contexts. It differs in this and many other ways from more recent theories or models of intonation, such as the autosegmental-metrical approach, which is now much more influential in the general description of intonation, not only for English but also for other languages (see Section 1.6). However, the traditional British school is still quite influential because of its long history and detailed analyses of the use of intonation, both in general phonetics and, especially, in more didactically oriented circles, such as teaching English as a foreign or second language (TEFL/TESL). It has special relevance for my investigation for two reasons: first, some of the later incarnations of this descriptive framework (Gimson 1989, Cruttenden 1997) make relatively explicit predictions about the prominence levels found in different positions of an utterance or phrase, through a hierarchical system of stress/accent levels, and some of these predictions contrast markedly with the most common predictions for Danish. These putative differences are a central part of this thesis. Second, the British tradition is by far the most influential description of English intonation in Danish books on English phonetics, and as such has had an enormous impact on how English intonation is taught in the Danish school system, from elementary school to University level (Davidsen-Nielsen 1984, Davidsen-Nielsen 1994, Livbjerg and Mees 1997, Mees and Collins 2002). This means that the above-mentioned differences in predicted prominence levels between English and Danish are a central part of the contrastive description of Danish and English intonation found in these books.

One of the earlier descriptions is Armstrong and Ward (1931). They set out to account for the 'simplest forms of intonation used in conversation and in the reading of narrative and descriptive prose' (Armstrong and Ward 1931: 1), and to this end use two types of 'tune', that is, two patterns of intonation which characterise an entire phrase (often an entire short utterance). Tune I is used in declarative statements, commands and questions other than 'yes/no-questions', and Tune II is used to signal less definiteness and in yes/no-questions. An example of a sentence with Tune I can be seen in Figure 1.1.

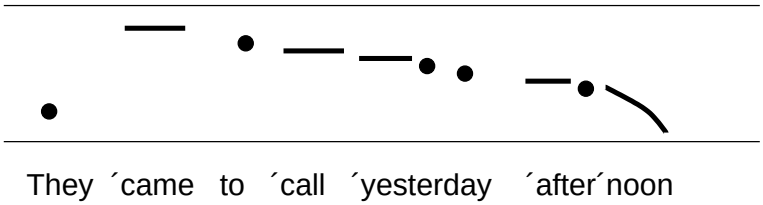


Figure 1.1. Example of sentence with Tune I. Adapted from Armstrong and Ward (1931: 4). Dots denote unstressed syllables, lines denote stressed syllables (which are also marked in the text by a ´).

The stressed syllables are pronounced on a gradually lower pitch through the sentence, and the final stressed syllable has a downward pitch glide – a fall. This

representation of the intonation of declarative sentences is common to almost all descriptions of English intonation, with some modifications and/or additions.

Armstrong and Ward distinguish two levels of stress – stressed and unstressed – and consider stress to be the 'breath force which we use in speaking' (Armstrong and Ward 1931: 3) to mark important words or ideas in a sentence. In other words, the same definition of stress as Jones (1918). They do claim, however, that other factors, mainly pitch changes, contribute to the perception of prominence, and note that it is sometimes difficult to determine the relative contributions of stress and pitch in this respect.

Armstrong and Ward claim not to have been influenced by earlier description of English intonation, including Palmer (1922), which appeared a few years earlier. In Palmer's analysis each sentence, or rather each 'Tone-Group', is divided into three sections, namely 'Head, Nucleus' and 'Tail', and he states that 'the conception of *Nucleus*, *Head* and *Tail* is my own' (Palmer 1922: vii). This type of analysis and the terms used in his book are found in most subsequent descriptions of English intonation and can be seen as a defining characteristic of what we call the 'British school of intonation analysis'. It is particularly noteworthy that the term *Nucleus* is used here for the first time, although Palmer owes a certain debt to the work by H. Klinghardt, and according to Cruttenden (1990) similar concepts can be traced back through H. Sweet and A. Bell to I. Watts in the 18th century and possibly even further back. In terms of definition, Palmer writes that 'Each Tone-Group contains a **Nucleus**, which is the stressed syllable of the most prominent word in the Tone-Group' (p. 7, emphasis given). There are two important things to note in this definition: (1), that the nucleus is expected to be present in each tone-group, although certain exceptions are noted elsewhere in Palmer's description, and (2), that it is defined with reference to prominence. Head and tail are the parts of the tone-group which precede and follow the nucleus, respectively, and are both optional elements. Except for the definition of the nucleus Palmer does not go into the issue of prominence or the relation between stress and intonation.

In Kingdon (1958a) and Kingdon (1958b) these issues are covered in some detail, and in general the distinction between stress and intonation as linguistic systems is clearer than in most previous work, although they interact in a rather complex way in some of his descriptions. *Stress* is defined as 'the force employed in uttering a syllable, giving it a certain degree of prominence', and *sentence stress* as 'the relative degree of force given to the various words in a sentence or utterance' (Kingdon 1958b: glossary), that is, stress is still seen as physical effort and unrelated to variations in pitch, although the interaction with this intonational parameter leads to the stipulation of several, albeit partly unspecified, levels of stress (e.g. Kingdon 1958a: 46).

Kingdon modifies Palmer's structural analysis and definition in a few ways: the Head is further divided into Prehead and Head (proper), where the Prehead is any unstressed syllable(s) that may occur before the first stressed syllable. Furthermore, the nucleus, or Nuclear Tone, is said to be 'associated with the last fully stressed

syllable of the group', in addition to being the most prominent syllable in the group (Kingdon 1958a: 6, 38).

O'Connor and Arnold (1973) do not refer directly to prominence in their definitions but distinguish between *accent* and *stress*, where accented words are those which are marked as particularly important in terms of meaning, while stress merely has relevance for the rhythmical aspects of an utterance (O'Connor and Arnold 1973: 7, 16). Accents always fall on stressed syllables, and can either be marked by pitch movements or by their position in the tone group in combination with stress. Nuclear syllables are defined with reference to their position as 'the stressed syllable of the last accented word', rather than to special prominence, but they are referred to as 'landmark[s] of the highest importance' which are 'made to stand out by a combination of stress and the pitch features of the nuclear tone' (O'Connor and Arnold 1973: 14-15). This means that O'Connor and Arnold distinguish at least three levels of stress/accent: unstressed, stressed but unaccented, and accented, and possibly with a fourth level for nuclear syllables. Which level, or type, is found on a particular syllable or word depends partly on where it is found in the tone group: as the last accented word (nuclear), within the head (accented) or outside the head (stressed, unaccented).

Other descriptions of English intonation are very explicit in setting up a four-level scale or hierarchy of stress and accent, namely Gimson (1989) and Cruttenden (1997). The latter is one of the most recent descriptions of English intonation according to the traditional British framework. These two works are very similar in their treatment of stress/accent levels, but the account below is based mainly on Cruttenden (1997). The four degrees of stress/accent are defined in Figure 1.2, reproduced from Cruttenden (1997)

(i) PRIMARY STRESS/ACCENT involving the principal pitch prominence, i.e. the NUCLEUS.
(ii) SECONDARY STRESS/ACCENT involving a subsidiary pitch prominence in an intonation-group, i.e. a non-nuclear pitch accent.
(iii) TERTIARY STRESS involving a prominence produced principally by length and/or loudness. (This is not referred to as 'tertiary accent' because the term 'accent' is reserved for pitch prominences.)
(iv) UNSTRESSED
(The term 'unaccented' covers (iii) and (iv).)

Figure 1.2. The stress/accent hierarchy in the British school of intonation analysis as represented in Cruttenden (1997: 44).

The three degrees of stress above the level of unstressed are defined with reference to how they are realised phonetically. Primary and secondary *stress/accent* are made prominent by means of *pitch changes*; it is not made clear here how these two accents differ, but explanations and exemplifications later in the book point to a definition similar to the one in Gimson (1989), namely that primary accent is 'signalled by means of a change of pitch *direction*' and secondary accent is 'signalled by means of a change in pitch *level*' (Gimson 1989: 270). Tertiary *stress* is prominence by means of phonetic properties other than pitch. One important *terminological* difference between Gimson (1989) and Cruttenden (1997) should be mentioned here. Gimson's system does not include 'tertiary stress' – only primary accent, secondary accent and unaccented. But since his secondary accent is divided into two types, with and without pitch prominence, the system is essentially the same as Cruttenden's.

The various degrees of stress/accent are associated with syllables or words marking certain intonational events or positions in an utterance or tone group. The nucleus, or primary accent, is the most prominent syllable and is often found on the stressed syllable of the last lexical word of a tone group. It is always, and by definition, the last syllable with pitch prominence in the intonation unit. Secondary accents are found in the head of the tone group – the first such accent is sometimes referred to as the *onset*[1], but changes in pitch level later in a head can also lead to the perception of a secondary accent. Tertiary stresses and completely unstressed syllables can be found in any position in the tone group. Some of these predictions about stress level will be tested in the experiments reported later in this thesis.

The fact that the nucleus is by definition the last syllable with pitch prominence in the intonation unit points to a sometimes unfortunate interdependence between nuclei and phrase boundaries: if two syllables in an utterance are perceived as stressed and are associated with pitch movements or a change in pitch level they must be regarded as nuclei, and by definition they must be separated by a boundary. This interdependence can be problematic in cases where other phonetic properties do not indicate a boundary – a problem which is treated in more detail in Section 4.2.1.

The issues of emphasis, focus and contrast are not covered in the hierarchy above, although some of these concepts received much attention in some of the earlier work on English intonation in the British tradition. Coleman (1914) distinguished two types of emphasis. The first concerned giving emphasis with the purpose of drawing attention to specific words, for example in connection with an explicit contrast between two items, as in the example

> *You may call it* DARK BLUE. *I should say it was* BLACK. (Coleman 1914:8)

[1] It is more often referred to as the *head*, but since *head* is also (and more frequently) used to denote the part of the intonation unit from the first stressed syllable to just before the nucleus, the term *onset* will be used in the following to indicate the first stressed syllable of the intonation unit.

which he refers to as *Prominence*. Coleman's *Prominence* thus corresponds to what many would call *narrow focus* today, and the above example is a special type of narrow focus, namely contrastive focus. The other type of emphasis, which is less relevant for the present investigation, is *Intensity*, which is emphasis to add an effect to an (already established) meaning, as in

> *I* INSIST *upon it. It is* ESSENTIAL. (p. 8)

Similar concepts are found in Armstrong and Ward (1931) and Kingdon (1958a), and it is noted in all these works that pitch plays a significant role in signalling these functions, through raised pitch on the emphasised item or an increase in pitch range. Kingdon also notes the possibility of emphasising an item by 'reduc[ing] the emphasis on the rest of the utterance containing it, so that it stands out by contrast' (Kingdon 1958a: 38). Kingdon also describes how a word can be emphasised by changing the position of the nuclear tone from its default location at the end of the utterance to the emphasised word. This description of emphasis (the type referred to as *Prominence* by Coleman) as a question of nucleus placement is very common today, e.g. in Cruttenden (1997), and is of course directly linked to the issue of focus (and contrast).

1.6 Stress and the autosegmental-metrical approach

The Autosegmental-Metrical (AM) approach to intonation analysis was started by PhD theses by Liberman (1979) and Pierrehumbert (1980). It describes the intonation contour of an utterance as a series of local events, either pitch accents or edge tones, which are associated with prominent syllables and boundaries between structural domains respectively. It is in this respect very different from the British tradition which analyses intonation contours into tone groups, or intonation units, with a specific tune.

The AM approach has gained enormous support since it was first applied to the intonation of American English (Pierrehumbert 1980), and the intonation systems of many languages have now been described within this general framework. However, this approach has so far had no significant influence on basic textbooks on English intonation, including those that are used in the Danish system of education (see above). Therefore, it will be given only a cursory treatment here.

In his account of the AM theory Ladd (1996) points out that stress and accent should be seen as two entirely separate systems, but unlike Bolinger (1958), who saw stress as a lexical specification (a propensity for accent) and (pitch) accent as the manifestation of prominence in the utterance, Ladd claims that both stress and accent have reality at utterance level. The exact degree of stress, or prominence, is determined by the hierarchical metrical structure of the utterance through the association of either strong or weak nodes with constituents of different domain sizes, such as syllables, words or phrases (Ladd 1996: 59). Pitch accents belong not to the system of prominence but the system of intonation. They are associated with metrically prominent syllables and are often aligned with these in the utterance, in terms of acoustic features – especially F_0. As such they serve as cues to the location of the

prominent syllables, although they are not seen as the direct cause of this prominence. Ladd states that 'stress [...] might be glossed as "acoustic salience": it is a complex of properties that can be related to greater force of articulation, including increased intensity and duration, and shallower spectral tilt' (Ladd 1996: 58). However, pitch accents certainly add to the perception of prominence in the sense that the scaling of F_0 peaks associated with pitch accents is directly correlated with the perception of degree of prominence (Liberman and Pierrehumbert 1984, Gussenhoven and Rietveld 1988, Terken 1991).

Perhaps one of the most notable manifestations of the AM theory is the development of a notational system for transcribing intonation, initially for English but now also for other languages such as German, Japanese, Korean, Greek, Spanish and Dutch, namely the ToBI (Tones and Break Indices) framework (Silverman *et al.* 1992, Pitrelli *et al.* 1994). In the guidelines for the English version of this system the connection between prominence and pitch accents is described as somewhat more direct than in Ladd (1996) and a further contrast is pointed out: 'Above the level of contrast between pitch-accented versus unaccented words, native speakers of English can distinguish another level of stress contrast, that between the last accented word of a phrase and any preceding accent' (Beckman and Elam 1997: 11). That is, not only do pitch accents signal stress (distinctions) but the final accent, or nucleus, is more prominent than other accents in the phrase, just as in the traditional British descriptions. The prominence relations may still stem from underlying metrical structure, of course, but the relation between prominence and pitch accents seems very clear. Furthermore, the connection between prominence level and position in the phrase, with regard to the identification of a nuclear accent points to a connection between phrasing and prominence: the presence of a boundary affects the status of the preceding accent (as nuclear), and conversely we might expect the perception of 'extra' prominence to lead to the perception of a boundary. Again, this is very similar to the British framework.

1.7 Stress in Danish

The manifestation of stress in Danish differs in a few important ways from the manifestation of stress in English, but the acoustic parameters which are associated with stress are the same, namely F_0, duration and intensity, in addition to vowel quality (spectral tilt or emphasis has not been investigated for Danish). Following the findings of Berinstein (1979) that any parameter which is also used for phonemic distinctions in a language will be moved down the hierarchy (see also above), Thorsen (1980) comments that in Danish, which does have phonemic vowel length, the expected hierarchy should then be F_0, intensity, duration, but she questions whether intensity really comes second. Danish has a high degree of reduction in unstressed syllables and Thorsen suggests that the hierarchy is rather: 'F_0, vowel duration and quality, intensity' (Thorsen 1980: 124), although the subject was (and is) still open to experimental verification.

Fischer-Jørgensen (1984) investigated the acoustic manifestation of main, or primary, stress as well as secondary degrees of stress in compounds, and found that syllables with main stress differ from syllables with secondary or no stress by F_0 and duration. Intensity differences were found between main stress and no stress, but not (or hardly) between main stress and secondary stress, while syllables with secondary stress behave like unstressed syllables with regard to F_0 (Fischer-Jørgensen 1984). The connection between stress and F_0 has been studied extensively by Thorsen, later Grønnum, as part of her research on Danish intonation (Thorsen 1978, Thorsen 1980, Thorsen 1982, Grønnum 1992). The exact F_0 movements that signal stress vary across Danish regional varieties, but in Standard Copenhagen Danish (main) stress is characterised by an upward movement in F_0 from the stressed syllable to the following unstressed (or 'post-tonic') syllable, with gradually descending F_0 in any following unstressed syllables. The F_0 movement, if any, in the stressed syllable is typically falling, but may also be rising or falling-rising depending on certain factors such as the position in the utterance and segmental composition (Thorsen 1982), but this variability is not linked to any syntactic or semantic/pragmatic factors such as is postulated for English (Kingdon 1958a, O'Connor and Arnold 1973).

One crucial difference between Danish and English is the lack of an obligatory nuclear accent, or sentence accent. In neutral utterances in which no word has been emphasised for semantic or pragmatic reasons all syllables with main stress will normally be perceived as equally prominent without leading to a sense that the utterance is unnatural or unfinished. Nor is there any additional tonal movement associated with the last stressed syllable (in fact, because of the gradually decreasing F_0 range throughout the utterance the opposite can be said to be true). Little is known about minor variations in the perceived strengths of syllables with main stress over the course of the utterance, but Grønnum states that variations in the degree of perceived prominence are typically associated with minor variations in the rise of F_0 of around 1–2 semitones (Grønnum 1995, Grønnum 2003). Emphasis for contrast is achieved by reducing the F_0 movement in the stress groups surrounding the emphasised item, generally, but not necessarily, accompanied by raising the F_0 level and exaggerating the F_0 movement in the emphasised syllable (Thorsen 1980: 171).

1.8 Comments on terminology and definitions

The previous sections focused mostly on the different views on what constitutes stress and especially how it is manifested and perceived, and did not deal in depth with the very considerable terminological differences which exist in the literature, both historically and, to a lesser extent, in the current literature. These differences do not only concern the choice of words to describe certain phenomena, but are rooted in different perceptions of the linguistic functions and relevant domains of these phenomena. Consequently, it can be very difficult to compare analyses and results across such differences without access to (relevant samples of) the material which underlies these analyses. Rather than trying to tie up these accounts into a unified and coherent whole I will here give a brief account of what I believe to be common

perceptions of the terms *stress, accent* and *prominence*. This cannot be said to represent a consensus view – merely a view which is found frequently in the current literature.

Stress is used in two very different ways. In the first sense, it is a lexically specified property of words whose domain is the syllable. That is, for each word (at least) one syllable is marked as having the possibility of being made prominent in an utterance. This is the basis for minimal stress pairs such as *pérmit* (noun) – *permít* (verb), which have lexical stress on the first and second syllable respectively. The words can then be distinguished purely by their stress, if and only if the stress is actualised in an utterance. The second use of the word stress is as actualised utterance level prominence. In part, this is what makes us able to distinguish minimal stress pairs like the above example, but it can be used in a more general sense to indicate the prominence of words which are marked in the utterance as particularly important. This last usage is considerably less common today than it was up until the 1960s, but it can still be found in various research papers (Silipo and Greenberg 1999, Silipo and Greenberg 2000), and is the normal usage in descriptions of Danish. However, actualised utterance level prominence marking important words is now more commonly referred to as *accent*. Bolinger (1958) abandons the traditional definition of stress and suggests using this term for the lexical specification only and to refer to utterance level prominence as (pitch) accent and states that 'one possible kind of stress is POTENTIAL FOR PITCH ACCENT' (Bolinger 1958: 149). To Bolinger the defining feature of a pitch accent was, as the name suggests, pitch obtrusions and this established a very direct link between pitch prominence and the 'semantic peaks' of an utterance (what others have called the important words), and other acoustic features which are known to be associated with his 'accent', were considered secondary. However, although Bolinger's use of the term pitch accent has been extremely influential, many, or most, theoretical frameworks differ from this relatively simple equation. Both the traditional British school of intonation and the AM theory, as well as many other descriptions, operate with two levels or categories at the utterance level, namely both stress and accent, albeit in somewhat different ways. The term accent normally does refer to pitch, but where the British school (as represented by Gimson 1989 and Cruttenden 1997) operate with a hierarchy of stress/accent with pitch changes marking the highest degrees of prominence, the AM theory claims that stress and accent are two different systems, even if they are closely linked (Ladd 1996). The term *prominence* is the least controversial of the three and generally refers to the degree to which something stands out from its surroundings. It may be used about specific properties, such as pitch prominence, which indicates deviation in pitch, or more generally, as perceived prominence, about the overall degree of emphasis (or de-emphasis) of a certain item.

The investigation reported in this thesis is concerned mainly with variations in perceived prominence (and to some extent the associated acoustic cues) in both neutral utterances and utterances with a specific focus. It is my contention that such variations play a large role in the prosodic organisation of utterances and affect the way in which we interpret information structure (loosely defined) regardless of the

manner in which this prominence is achieved (but in close interaction with other prosodic parameters such as intonation and phrasing). And since I am not convinced that prominence which is brought about by pitch obtrusion, or variation in F_0, is categorically different from prominence by other means I will not use the term accent to denote prominent syllables or words in general; the term is too closely associated with this specific type of prominence, and to some extent also with specific theories of intonation, such as the AM theory. Instead, I will refer to utterance level prominence as *stress*, and in many cases it will in fact be synonymous with perceived prominence. When a more specific use of the term is needed to refer to linguistic organisation (lexical specification or stress levels) this will be made explicit in the text. I have, however, occasionally used the term accent when discussing theories or articles which have used this term, and in addition the term 'focal accent' is used to refer to the location of an item which is emphasised due to semantic focus or explicit contrast.

Although the terms stress and prominence are occasionally used more or less synonymously it should be pointed out that the issue of prominence scales, as described in Section 1.4.4, is of course different from and formally independent of linguistic systems of stress levels (or stress/accent levels). Stress levels are normally defined with reference to lexico-syntactic properties and as such characterise the internal organisation and rhythmic pattern of a word (simplex or compound) or certain phrase types with obligatory stress reduction. These stress levels represent linguistic *categories* which are assumed to distinguish meaning at the word (or short phrase) level, and any syllable will carry either primary stress, secondary stress or no stress (or any other level which may be defined in a particular system), but cannot have some intermediate level between these. The same holds true for the four-level stress/accent hierarchy of the British school of intonation analysis (Gimson 1989, Cruttenden 1997), although the levels in that hierarchy are defined by the organisation or internal structure of the intonation unit and in some cases the relative prominence of the syllables (do they have tonal prominence or not). Prominence on the other hand is a matter of more or less – it is 'continuously variable' – and there are, in principle but probably not in terms of cognitive reality, an infinite number of prominence 'levels.' The formal independence between stress levels and perceived prominence does not mean that there is no connection between the two. It must be assumed that the linguistic stress levels are reflected in the perceived prominence in such a way that *the relations in perceived prominence level between any two syllables/words within the same relevant domain should never cross the linguistically defined stress levels.* In other words, a 'secondary stress' should never be perceived as more prominent than a 'primary stress' within the same domain. I hesitate to define the 'relevant domain' more precisely at this point, but a good working definition might be the intonation unit or intonation phrase. Any deviations in the connection between stress and prominence as defined above should be reflected in the organisation of larger domains, for example the subordination of phrases to indicate parenthetical information or the like. A similar point is made by Pierrehumbert about the connection

between metrically determined prominence and the (perceived) prominence of associated pitch accents:

> The term 'prominence' will be used to refer to the aggregate of metrical strength and emphasis, as it pertains to the control of tonal values. We will assume that each pitch accent has an associated value, that prominence is continuously variable, and that the prominence of a metrically stronger accent is at least as great as that of a weaker accent, though not necessarily greater (Pierrehumbert 1980: 40).

If systematic violations of these expectations are found, not only with regard to the AM theory, but also for the British system or other systems of stress/accent levels, then it is an indication that the theory needs to be amended.

CHAPTER 2

An investigation of prominence – collecting data

2.1 Introduction

In the Introduction it was stated that the main objective of this thesis is to examine prominence relations in Standard Southern British English. The original intention was to provide an account of the acoustic correlates of prominence, with special attention to details in the timing of F_0 movements relative to the stressed syllable, and the role of spectral balance in marking various levels of stress or accent. The issue of perceived prominence was intended to play a relatively marginal role, mostly as a confirmation of the prominence assignments made by the author in the analysis of the recorded utterances. The results of this initial perceptual experiment suggested that it would be worthwhile to conduct a more thorough investigation into the perceived prominence level of the words in the material, so this perceptual part of the study was expanded and ultimately replaced the acoustic correlates as the core of the project. The data material that was used in the perceptual study was therefore collected with the original purpose of doing an acoustic analysis, which has had a significant effect on the choice of speech material and on the recording procedures; it is quite possible that the data collection procedure would have been different if the primary objective had been perceived prominence relations from the outset. This said, it should be pointed out that it was the results from exactly the type of material used which led to findings which prompted further investigations of perceived prominence. These findings might not have been apparent if the material had not been so carefully controlled for certain types of variation as had been deemed necessary for an acoustic investigation.

The purpose of the investigation as it is presented in the remainder of the thesis is then to describe the variation in perceived prominence level of words in short English utterances, both semantically and pragmatically neutral utterances and utterances in which a specific item has been emphasised or foregrounded due to semantic focus or contrast. In lieu of a proper acoustic analysis the utterances will be *characterised* in terms of the acoustic properties F_0 variation and duration.

The collection of data – choice of material and speakers etc. – is presented in the next section, but some of the more detailed procedures which are only, or mostly, relevant for an acoustic analysis have been relegated to Appendix A, Section A.2.

2.2 Data collection

The procedure chosen for the collection of data was influenced by the original purpose of conducting acoustic analyses, which meant that the material had to be produced and collected in a relatively controlled setting using high quality equipment.

The choice of data material was a matter of some concern. There are two common types of material: (1) spontaneous speech, or the often used variant 'unscripted speech', or (2) carefully constructed sentences read aloud under laboratory conditions and therefore often called 'laboratory speech'. The first option seems to be by far the most common one today. When examining journal articles and conference papers there are numerous references to analyses of 'unscripted speech', which usually means recordings of a kind of 'map task' activity (where one speaker gives instructions about a route on a map to another speaker with the same map without the route printed on it, see for example http://www.hcrc.ed.ac.uk/dialogue/maptask.html), or of recorded dialogues collected in a larger speech corpus. This constitutes a significant change from just ten years ago, when laboratory speech was still used frequently. There may be many reasons for this change: the tools for constructing large speech corpora have improved; the research may be meeting requirements from the speech technology industry who use the large corpora to construct more reliable or efficient applications in speech recognition and/or synthesis; and naturally the very likely correct assumption that analyses of (semi-)spontaneous speech are a better reflection of the way people produce speech in real-life situations. However, this type of material is not well suited for an investigation which originally intended to describe minute details in the variations of acoustic features under different contextual conditions. For this it is necessary that all (or most) other factors are kept constant. There are many known, and perhaps more unknown, interactions between pitch, duration and intensity and they are likely to all vary with position in the sentence, total sentence length and other factors. It was therefore decided to use constructed sentences which are read aloud in a studio, and with variation in only the parameters under investigation. For one of the original central acoustic cues – intensity – this procedure is an absolute prerequisite, since the recording level needs to be calibrated and the distance between microphone and speaker must be constant, but there are advantages even for the other cues.

If the original purpose had been to examine perceived prominence it is quite possible that some type of spontaneous speech would have been chosen instead, in order to obtain material with a richer variation in prominence levels and where this variation is used to signal pragmatic, or discourse related, information, but as I will argue below, such a choice might in fact have disguised the regular variation in prominence level over the course of an utterance which became apparent from the more restricted, or controlled, material which was actually used.

2.2.1 Text material

The reading material was kept as small as possible, since several repetitions of each sentence were needed (for quantitative analyses of the acoustic measurements) as

well as more than one version of many of them, with a different focus structure, or foregrounding/backgrounding, in each of them. There were 11 semantically different sentences, three of which had two variants, plus five 'filler' sentences which were not analysed. The 11 sentences (14 if the variants are counted separately) which thus make up the material proper are listed below, including the questions that were used to elicit different foregrounding strategies.

Sentence material

(The abbreviations in parenthesis after each sentence are used to identify the sentences in the presentation and discussion of results below – that is, as shorthand names. In some cases only the part of the sentence which has been underlined is used in the analyses.)

1. Paul sings. (*ps*)
   Questions:
   Who?
   He what?
2. Bill struck Ann. (*bsa*)
   Questions:
   Who did?
   He did what to her?
   Who did he strike?
3. Jane kissed Frank tenderly. (*jkft*)
   Questions:
   Who did?
   She hit him?
   She kissed Fred?
   How (did she kiss him)?
4. The party was cancelled. (*pc*)
   Questions:
   The match was?
   It was postponed?
5. The cook was smelling the soup. (*css*)
   Questions:
   Who?
   He was eating it?
   He was smelling the wine?
6. Sheila examined the patient carefully. (*sepc*)
   Questions:

Who did?
She instructed the patient?
Who did she examine?
She examined him quickly?

7a. A widespread shortage of aluminium has forced many European countries to import kitchen sinks from Denmark.
We don't know about the Italians, but we do know that the Germans import sinks. (*imp_vb*)

7b. It is a well-known fact that, in times of financial crisis, many countries tend to reduce the amount of net import of goods.
However, in such situations we don't know about the Italians' import, but we do know that the Germans' import sinks. (*imp_sb*)

8. The Germans' import of sinks from Denmark has been a boost to the Danish economy. (*tgios*)

9. The Germans import their sinks from Denmark. (*gitsd*)

10. Is Peter a doctor in Paris? (*pdp*)

10b. (Same as 10. but in the context:
A: Jane is a doctor in Paris, and so is Arthur. Do you know anybody else?
B: Is Peter a doctor in Paris?)

11. Did Stalin insist on an equal distribution of wealth? (*dsi*)

11b. (Same as 11. but in the context:
It was an important idea in the early days of the Soviet Union that the nation's wealth should be distributed equally among the people. Many of the Soviet presidents adhered strongly to this idea.
Did Stalin insist on an equal distribution of wealth?)

Default reading

The reading material has been constructed in a way which addresses the different issues that were mentioned in Section 2.1 using only 11 (14) sentences. In order to be able to do so one assumption had to be made: namely that the manner in which these sentences would be read could be predicted and controlled. More specifically it was assumed that each sentence has a *default reading*, a particular way in which that sentence is read when there is no semantic or pragmatic context which yields a more specific interpretation. The notion of default reading applies in particular to accentuation, that is, which words are stressed and which are not. For example, in the sentence:

Jane kissed Frank tenderly

the default reading was expected to have four stressed syllables, one in each of the four words (all are lexical, or content, words).

Along the same lines, but somewhat differently, I also expected to be able to elicit a focal accent on a particular word by asking a question relative to that word, or constituent. While such an utterance with a predictable accent pattern might not be called a default reading it was still an important premise for the experiment that these predictions could be made confidently. It was clear from the initial, informal analyses of the material and confirmed by the perception experiments reported below that these predictions were in fact borne out to a very large degree.

The 11 (or 14) sentences serve slightly different purposes in the investigation, and as such fall into three groups.

Group 1: sentences 1–6.

Sentences 1–6 may be considered the 'core' of the experiment, since it is through the variations found in the different renditions of these that most of the issues under investigation will be approached. They fall into two subsets, 1–3 and 4–6, the three sentences in each set having two, three and four lexical items respectively (resulting, so it is predicted, in two, three and four stressed words). They can therefore provide information on the manifestation of stress in sentences of different length, and focal (or emphatic) accents in different positions in the sentence: initial, medial and final. Sentences 1–3 contain only monosyllabic lexical items (except *jkft*, where the last word 'tenderly' has unstressed syllable *following* the stressed syllable), so all the syllables which are expected to be stressed are immediately adajcent, whereas sentences 4–6 have 1–2 intervening unstressed syllables between each stressed syllable, making it possible to evaluate the influence that unstressed material has on particularly F_0 configurations and duration.

The questions which are used to elicit repetitions with added emphasis on a particular item are of course rather important. There are several ways of doing this: one way is to replace the item one wishes to have focused with an interrogative pronoun; for example, in the sentence *Paul sings* one can ask 'Who?' or 'He what?'. Alternatively the item of interest can be replaced with a different lexical item, which then needs to be 'corrected' in the response, as in *The party was cancelled* followed by 'The match was?' or 'It was postponed?'. The two methods result in what has often been distinguished as different types of focus, and while they both can be said to constitute *narrow focus*, following the terminology in Ladd (1996), this term is more often used for the first method only, see for example Sityaev and House (2003), while the second type is recognised as a special kind of focus, often called *contrastive focus*, because there is an explicit contrast between two lexical items (although in two separate utterances) (Cohan 2000). Both methods were used; I generally preferred the first method, especially if the question could be kept short and maybe contain only the interrogative word which referred to the item under investigation. Furthermore, I did not want to repeat any of the words from the original sentence, to avoid any danger of 'mimicry', where the speakers, perhaps unconsciously, adopt elements from

my pronunciation in their reply. Therefore names and other lexical items were replaced with pronouns ('he, she, it') wherever possible.

With some of the sentences it was difficult to follow these general rules without creating unduly complex questions. Consider for example the sentence *Sheila examined the patient carefully*. The question 'She did what?' would most likely lead to a broader focus than required, namely on the whole predicate *examined the patient (carefully)*, and restricting the scope of the focus would be difficult, so in this case, and others, it was better to use the second method 'She *instructed* the patient?'. I only included material that was necessary to get the desired response, cf. the absence of the word *carefully* in the previous example. I also provided prosodic information about which word was to be focused by using a focal accent on the relevant constituent when asking the question.

The variation in type of focus which was introduced by using two different methods to elicit utterances where one item is emphasised, or foregrounded, was not intentional, in the sense that it was to be included as a parameter in the experiments, but rather an oversight. This might constitute a problem since some studies, for example Sityaev and House (2003), have shown that the two types of focus are not necessarily realised in the same manner. But the two methods do not seem to have led to systematically different realisations of focus in my material. At least, it was not immediately obvious by listening to the different utterances, and an analysis of the prominence levels of focused items and the non-focal items in the same utterances (from the perceptual experiments reported below) showed no clear difference between the two conditions. There was no statistical difference between the prominence level of the focused items ($F(1,1550) = 0.312$, $p = 0.58$), while the non-focal items were deemed slightly less prominent in utterances with (semantic) focus than in utterances with contrastive focus ($F(1,3326) = 15.240$, $p < 0.01$). This difference is in the opposite direction of what might be expected and can be caused by variation in other parameters with which focus type covaries in the material (pre- versus post-focal position etc.). It can be concluded that there does not seem to be an (interfering) effect of focus type. The reason for this may be found in the larger context of the elicitation procedure. Speakers would read first a neutral version of an utterance, then I would ask the relevant question, which they then answered with an appropriate indication of focus. Regardless of type of question this can be seen as a *correction* procedure: my question to them indicates that I have misunderstood, or not understood, part of the previous utterance and they then supply the missing or correct information. It is possible that it is this type of discoursal function of correcting which is responsible for the particular, and seemingly identical, prosodic strategy used by the speakers in the two semantically/formally different types of focus condition.

Group 2: sentences 7–9.

These sentences were originally included to be able to test if the verb versus noun contrast in pairs such as *import* (vb.) – *import* (n.) is maintained, that is, present, in

post-focal position (7a, 7b) and, for comparison, in normal, non-nuclear, stressed position (8, 9). The analyses of this phenomenon are not included in the thesis, but sentences 8 and 9 were used in the perceptual experiments as examples of neutral utterances.

It was brought to my attention during the recording sessions that sentence 7b is not strictly grammatical. The noun *import* either requires a modifier (… *of goods*), or must have the plural form *imports*. Several of the six speakers commented on this, but only one speaker seemed troubled by it. All the speakers were experienced readers and I do not believe that it had a significant effect on the way this sentence was read.

Group 3: sentences 10–11.
All the sentences in groups 1 and 2 are terminal declaratives, but two interrogatives (yes/no questions) were included in order to obtain a different intonation pattern. Obviously, a rising pattern is not in any way obligatory with yes/no interrogatives, or even that common in ordinary speech, but it was still expected that some rises could be elicited in this way without having to instruct the speakers to use a rising intonation. The b-versions of these sentences have a pre-context which promotes the placement of a focal accent on the first lexical item in the sentence.

Reading lists
All the combinations of sentences plus focus-related questions (and filler sentences) amounted to 32 sentence items in total. An *item* could be the sentence 'Paul sings' read once, followed by my asking 'who?' and then the repetition of 'Paul sings' as an answer to that question. Eight randomisations of the 32 items were made, with one restriction: sequences of the same sentence were not allowed and where this occurred as a result of the randomisation process the sentences were manually rearranged.

The speakers
Six speakers recorded the sentence material – a number which strikes a good balance between the statistical, or other quantitative, requirements and the time available for analyses.

All the speakers have a southern British linguistic background, although in slightly different ways. One person was born in Wales, but moved to England at age 6, another had lived in Sheffield during her secondary school years. But they all described their accent as either 'RP' or 'RP-like', especially when reading aloud in a sound studio. In other words, they generally represent Southern British English (SBE) with a few slight variations – much as you would find in a 'real-life' situation – but not RP in the traditional narrow sense of public school pronunciation.

There were three male and three female speakers, between the ages of 29 and 59. See Appendix A, Section A.1 for a more detailed description.

2.2.2 Recordings

The recordings were made in a soundproof studio at the University of Edinburgh in Edinburgh, Scotland. I was in the studio with the speaker, while a technician took care of all the technical details (described below) following my instructions. The test sentences were presented on separate cards (one sentence item on each card). Although the speakers were instructed not to turn the card until they had read an item completely, there are some utterances which have slightly overlapping paper shuffling noise. It took about five minutes to read one session/randomisation, and the speakers were allowed a short break while the technician made preparations for the new session. The total recording time was 45–50 minutes, including a small 'map task' like activity (giving instructions about assembling a small jigsaw puzzle) which was not used for this study.

Recording setup

The recording setup included two microphones; one 'shotgun' type directional microphone placed approximately one meter from the speakers and one 'clip-on' microphone fastened to a headband on a metal extension which placed it approximately 13 centimetres from the speaker's mouth. This relatively intricate system was motivated by the need for controlling the distance to the microphone for the intensity measurement. More details about the recording setup and the calibration techniques can be found in Appendix A, Section A.2.

2.3 Selection of utterances

The recorded utterances were submitted to a selection process to determine which were to be included in the subsequent analysis, especially with regard to the quantitative acoustic analyses or descriptions. This was necessary for two reasons: (1) quantitative analyses can only be performed on utterances which represent the same structure, that is, which are felt to be 'perceptually equivalent', and which furthermore achieve this equivalence in (roughly) the same manner acoustically, and (2) some of the sentences were represented by more token utterances than were needed for analysis, namely the neutral, context-free version of the utterances in Group 1, for which a neutral version was recorded for every version with a focal accent, that is, up to four times as many as required (with four different placements of the focal accent in *jkft* and *sepc*).

'Sameness/perceptual equivalence'

The original intention of making detailed quantitative analyses of acoustic properties such as F_0, duration and intensity, as well as the less detailed but still quantitative description presented in Section 2.5, requires that the individual tokens can be considered repetitions of the same entity. One might argue that a series of readings of the same sentence by the same speaker by definition fulfils this requirement, but considering the many different production strategies a speaker can adopt for one semantic sentence in order to convey a particular meaning, or information structure,

this condition is not sufficient. This relates to, for example, foregrounding – backgrounding of certain parts of the sentences, choice of intonation (rising, falling, level) and other factors which involve exactly the parameters under investigation here. Therefore, only utterances which convey the same meaning in the same manner can be grouped together for further analysis. While it is straightforward to determine whether two utterances have the same syntactic structure, it is not as easy to determine whether they have the same *prosodic* structure, and the requirements for 'equality' or 'sameness' will have to be somewhat laxer than would be necessary for syntactic equality.

The IPO school of intonation research has developed a formal procedure for determining whether two instances or versions of an utterances can count as the 'same'. The procedure, and the definition of some of the concepts involved in this, varies somewhat in different works from the IPO school (de Pijper 1983, Cohen and 't Hart 1967, 't Hart and Cohen 1973, Willems 1982), but the idea is always to find the relevant pitch movements in an utterance by making a resynthesised, simplified version which is equivalent to the original. The procedure is summed up in 't Hart, Collier and Cohen (1991), who point out the distinction between perceptual *equality*, which requires that the simplified version be auditorily indistinguishable from the original, and perceptual *equivalence*, where the two versions need to be equal in terms of perception within a linguistic system. The definition is as follows:

> if for a speech utterance two different courses of F_0 are similar to such an extent that one is judged as a successful imitation of the other, we say that there is perceptual equivalence between the two ('t Hart, Collier and Cohen 1991: 47).

According to their description this evaluation must be performed by a native speaker of the language under investigation.

My approach is both less formal and less strict than the IPO approach as outlined here, but I have retained the description 'perceptual equivalence', because the underlying premise is the same: two utterances are perceptually equivalent if they can be regarded as two tokens of the same linguistic structure. My task was different, since I was not comparing simplified copies with complex originals, but rather several original readings with each other, but I was still making an evaluation of whether one utterance could be considered a successful imitation of another. The requirement that the evaluation be performed by a native speaker could not be met in this investigation, but as is also demonstrated by the results of my listening experiments below this need not be a major concern.

To sum up, the first part of the selection process consisted of an auditory evaluation of each sentence by one listener – a non-native speaker of English with near-native proficiency (the author). Only utterances which were felt to represent the same linguistic (including prosodic) structure – in other words were perceptually equivalent – were grouped together and submitted to the second part of the evaluation process.

Acoustic similarity

The utterances not only had to sound similar, it was also necessary that the perceptual equivalence was brought about by similar acoustic properties. Of course, if this requirement is too strictly handled one can only determine whether two or more utterances are similar *after* the acoustic analysis has been performed, which would obviously defeat the purpose. So this stage should rather be seen as a screening process which attempts to find obvious outliers and exclude them from the analysis, and as a help in selecting utterances from the groups that included more than the required eight utterances (of one sentence by one speaker). For all utterances time-aligned waveform, spectrogram and F_0 displays were printed in hard copy, which were used to determine whether the utterances could be viewed as 'the same', that is variations of the same general structure. Not all the acoustic properties that would be part of the subsequent analysis were included in the evaluation. Intensity was not considered at all, and duration only to a moderate extent. If there were obvious and salient differences in syllable duration, or in the duration of (minor) pauses, this could affect the selection, but these aspects were most often caught in the perception stage of the evaluation. So the emphasis was on similarity with regard to F_0. Although the selection was performed from a fairly simple, visual inspection of the F_0 traces, there was one slightly more formal criterion that was strictly adhered to. In order for a series of utterances to be grouped together it was necessary that the F_0 contour of each entire utterance could be described in a simplified form by the same number of turning points at the same approximate locations. See Figure 2.1 for an illustration. This is also necessary for the generation of average traces of multiple repetition such as those presented in Section 2.5. Sometimes minor deviations were accepted if the auditory impression did not indicate that the acoustic differences were perceptually relevant. In Figure 2.2 the middle part of the F_0 contour has a downward slope in the left-hand utterance and an upward slope in the right-hand utterance, but the difference was judged to be imperceptible.

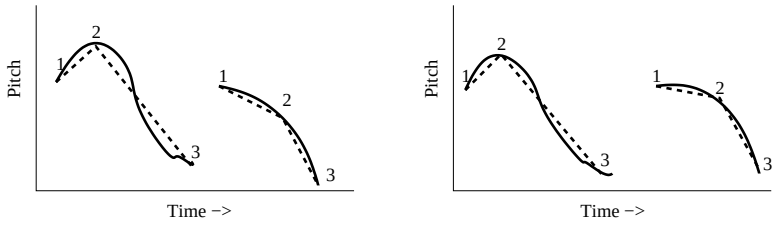


Figure 2.1. Illustration of two F_0 traces (solid lines) with an indication of the turning points which result in a simplified trace (dashed lines).

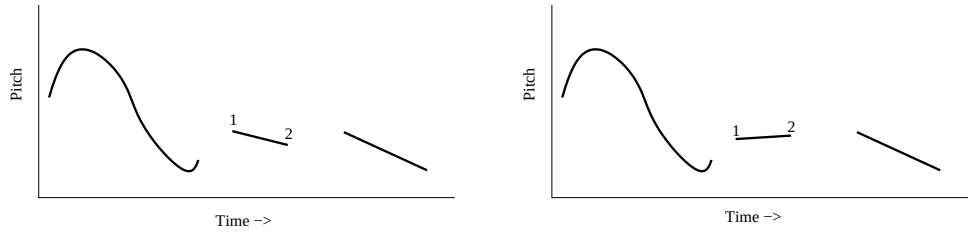


Figure 2.2. Minor deviations were accepted if they did not appear to be perceptually relevant, such as the slope of the middle parts of the F_0 traces above.

All sentences were read eight times (or more) by each speaker, and at least six repetitions were required in a group for my quantitative analyses (in order to be able to perform even basic statistical calculations of means and variance). Therefore, if fewer than six utterances in a group were judged to represent 'the same' utterance, the whole group was excluded. Based on this requirement and according to the evaluation process described above I had to leave out 2–3 groups of utterances for each speaker. For some of the neutral, default readings I had enough tokens of two different types of realisation of a sentence to include both types in the analysis. The total number of utterances selected for further analysis therefore varied between 30 and 32 groups for each speaker (183 total), each group containing six (or, rarely, five) to eight utterances, giving a grand total of 1351 utterances.

2.4 Segmentation

The next step in the collection of data was to mark up segmentation and measurement points in all utterances using signal analysis software (Xwaves/ESPS®), so that all measurement data could later be retrieved automatically using various scripts. This process turned out to be quite extensive and time-consuming, involving several steps: building a script to display the necessary views; inserting the segmentation points (approximately 40,000) in the label files; and writing or modifying scripts to extract information from the label files and ESPS data files. However, it is still much faster and, crucially, much more flexible than recording data manually on a sheet of paper. It is possible to make changes to segmentation criteria, adjusting the appropriate mark-up points and then run the scripts again with fairly little extra work, which indeed proved necessary. Below follows a brief outline of the process.

Segmentation tools

The segmentation process was run from a (shell) script which, for each utterance, presented several displays: pressure waveform, spectrogram, F_0 and intensity, all time-aligned. In addition, two label displays were attached to these; one for the segmentation points used to calculate duration, and one for F_0 measurement points. An example can be seen in Figure 2.3.

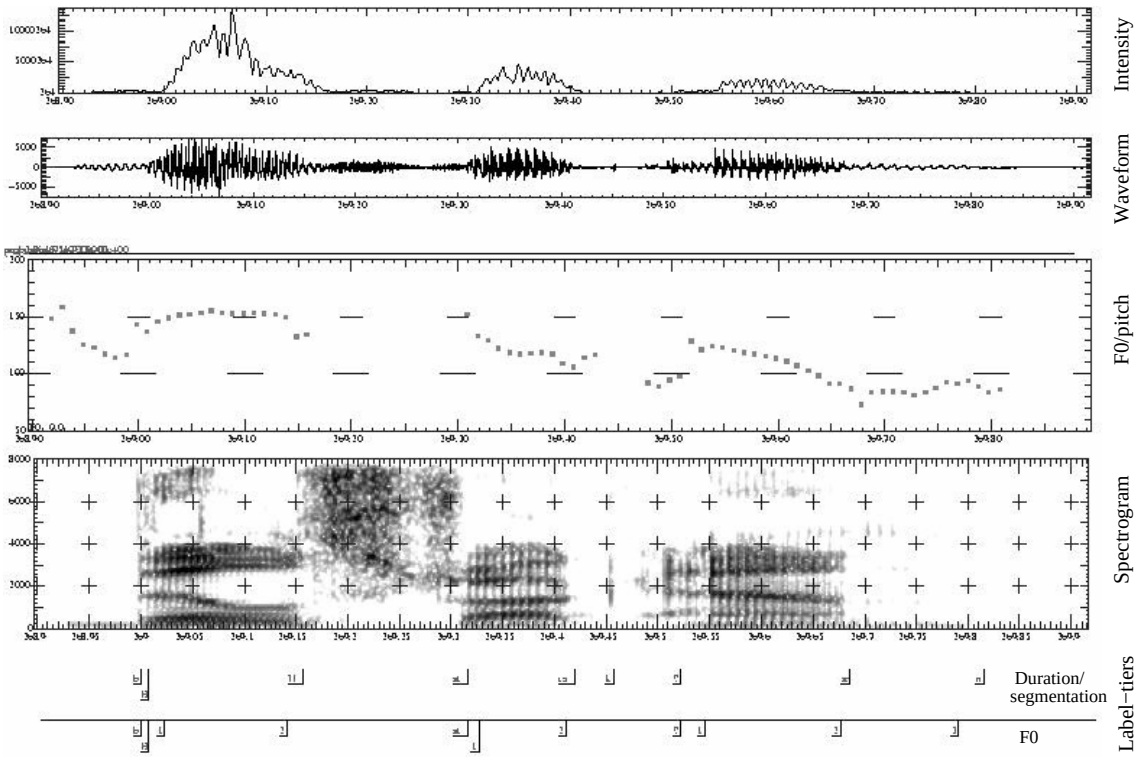


Figure 2.3. Segmentation and annotation view with various displays and label windows.

The measurement points necessary for calculating F_0 and duration are simply inserted in the appropriate label window with a symbol for each segment. The 'label' utility in Xwaves/ESPS® keeps the information from each label window in a plain text file which consists of a header and a number of lines containing the individual symbols and a time-stamp. It is thus very easy to calculate the duration of each segment by subtracting the value in the time-stamp from that of the following segment.

Measurement points

The segmentation in the duration tier of the labelling windows was, on the whole, done on a 'segment level' basis, that is, the phone level, as far as possible. In the case of, for example, V + [ɬ] or diphthong + [ə] where no segmentation boundary can be located the two segments were merged into one. Voiceless consonant clusters were not segmented into the individual phones, since such a segmentation was not deemed necessary for the analysis of either F_0 or duration (or intensity).

The F_0 tier was marked up in the following way: each stretch of voiced material, yielding an unbroken F_0 trace, was considered to be a whole, that is, a separate contour which had to be described with the exact same number of measurement points in all utterances in the same group (see above for further explanation, and Figure 2.1 for an example). These measurement points were coded in a way which allowed for an automatic re-tracing of the simplified contours. No indication was thus made about the alignment of the F_0 measurement points with the text, that is, the segmen-

tal string. Instead, this information must be retrieved by reference to the duration label tier.

Modification of the F_0 trace

Since the extraction of measurement data for F_0 is done automatically by a script which gets F_0 values at all specified points it is necessary that the F_0 analysis done by the signal analysis program (*get_f0* from the ESPS package) is 'correct' at the specified points. By 'correct' I mean a representation which is in agreement with other types of analysis, such as visual inspection of the pressure waveform or (narrowband) spectrogram, and at least not in total contradiction to some mode of auditory analysis, such as the formal approach of 'analytic listening' of the IPO school (Cohen and 't Hart 1967), or even simply the more informal impression of the investigator. But this is not always easy to achieve, since the F_0 analysis programs are far from perfect and the calculations may be disturbed by various irregularities in the speech signal. The result can be octave errors (upwards or downwards) or minor fluctuations which are not perceptually relevant. The problem was resolved partly by placing the measurement point in a section where the F_0 tracking seemed reasonable, in the case of minor and very local fluctuations (one glottal cycle), or by modifying the F_0 trace in the F_0 feature file, in the case of octave errors. This modification was done only after a careful inspection of the waveform and auditory confirmation that the analysis was erratic. With octave errors the F_0 curve was retraced at exactly double or half the analysed values, and in some instances a curve was smoothed by aligning single outliers with the surrounding F_0 values. Both octave errors and single outliers were quite common in my material, partly because of the widespread occurrence of 'creaky voice'. The total number of utterances in which some (usually minor) modification had to be performed was 398, or just under 30% of the utterances.

2.5 Acoustic characterisation of utterances

It was stated earlier in this chapter and in the Introduction that the detailed acoustic analyses which were originally planned had to be deferred until a later time. Much of the initial work in this process has already been performed, however, such as sorting and selecting utterances and marking them up for automated extraction of data (as described in the previous sections). The automated extraction process uses the timing information in the label files to gather information about F_0 at the specified points and to calculate segment durations. This information was used to create simplified averaged versions of all sentences (under different focus contexts) in the material for each speaker, which 'reduces' the 1351 utterances to around 180 (averaged) items. Simple visual inspection of this (fairly uniform) material provides a good impression of the overall tendencies in F_0 variation, so in the following sections selected tokens of the averaged versions are presented with comments on the general patterns. At the end of the chapter there is also a brief account of the variations in

duration in relation to position in the sentence and focus condition. An example of an averaged trace can be seen in Figure 2.4.

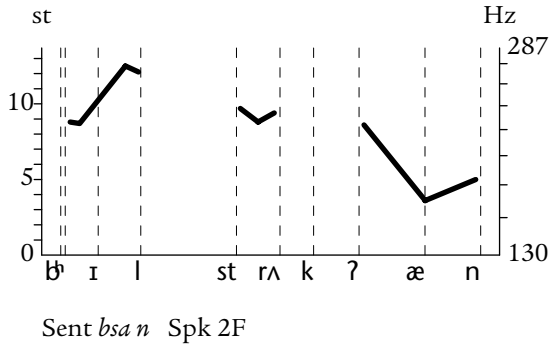


Figure 2.4. Average trace of sentence *bsa n* by speaker 2F. Both F_0 and duration values are averaged over all tokens of this sentence in the neutral, context-free version by speaker 2F. See further explanation of the diagram in the text.

F_0 is indicated on the y-axis: semitones at equidistant intervals on the left and Hz at intervals of 20 on the right, with an indication of the lower F_0 threshold for the speaker (set manually to the absolute lower threshold of the speaker) and the F_0 peak in the current utterances (set automatically). The solid lines represent the averaged F_0 traces, typically of 6–8 utterances. All diagrams in the following presentation have the same constant scale horizontally and vertically. Horizontally 1 cm corresponds to 200 ms. The right edge of a segment is marked by phonetic symbols; in case of very short segment durations there may be some overlap of symbols.[1] The sentence name and speaker identification are printed below the diagram.

2.5.1 F_0 range variability among speakers

There is considerable inter-speaker variability with regard to F_0 range – both in the neutral, context-free utterances and in utterances with emphasis (semantic or contrastive focus) on a specific word. Speaker 1F generally does not span more than 5–6 semitones within one utterance, sometimes slightly more in focused utterances, and her overall range in all utterances is around one octave. Speaker 4M frequently spans 1.5 octaves within one utterance, with an overall F_0 range of over two full octaves. The other four speakers are somewhere between these extremes: 5M is at the lower end and the remaining three roughly intermediate with typical F_0 ranges within one utterance of just under one octave. The extreme difference between speakers 1F and 4M is illustrated in Figure 2.5, showing average values for sentence *ps f2*.

[1] Note that the symbol [ʰ] is used to mark the end of the release burst of a stop consonant and any aspiration which might be present, sometimes collectively referred to as the 'open interval'. In other words, it marks the onset of the following vowel, whether the stop consonant is aspirated or not.

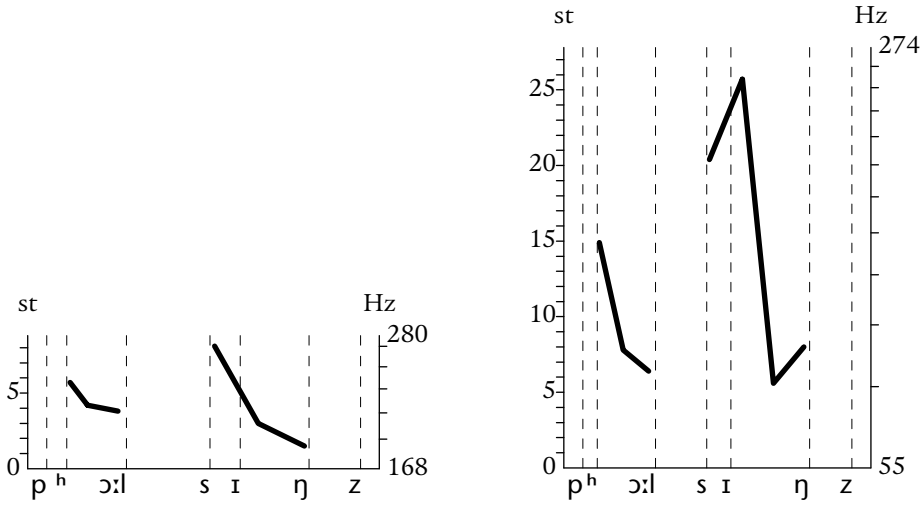


Figure 2.5. Example of F_0 range difference between the two speakers with the smallest and largest ranges in the material – here 7 (1F) and 21 (4M) semitones respectively. The utterance and focus context is the same: *ps f2*.

The large differences in F_0 ranges do not appear to be directly linked with perception of prominence level when judged utterance by utterance. In the prominence tests reported later in this thesis listener responses were coded on a scale from 0 (unstressed) to 3 (strong stress), and although speaker 4M with by far the largest F_0 range did achieve the highest overall rating (1.37, grand mean total), speaker 1F, with the smallest range was third, not far behind at 1.36.[2] The difference in overall prominence rating for all six speakers was small: from 1.30 to 1.37.

2.5.2 General observations about F_0

The utterances in the experiment are all produced in identical settings (language lab reading task), they are similar with regard to content, and are exposed to the same variation in focus (or information) structure. It is therefore not surprising that the material is quite homogeneous and that most utterances are variations on a common theme (or perhaps a few common themes). This makes it possible to account for common trends using a limited set of examples, but at the same time makes it difficult or impossible to generalise to 'English utterances' in a broader perspective. The presentation below will start with the utterances from Group 1 – the core material: first the neutral declarative sentences with and without stress clash; next the corresponding focused versions. Finally there are some examples of interrogative sentences from Group 3.

[2] The scores reported here are from Tests 1 and 2 pooled. See Chapters 4 and 5 for details.

2.5.2.1 Neutral utterances with stress clash

Two versions of the sentence *ps* with two lexical items are depicted in Figure 2.6. They represent the two types of realisation which were found in the material.

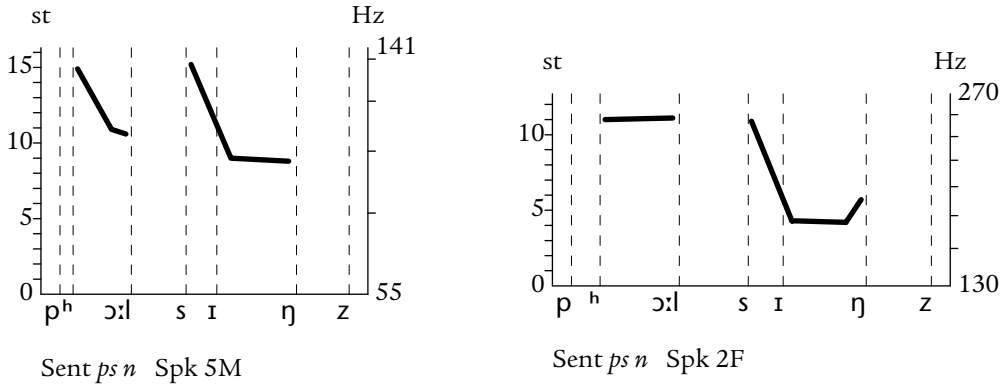


Figure 2.6. Sentence *ps n* by speakers 5M and 2F, representing two types of realisation: two falling F_0 trajectories, or level plus falling.

The type on the left-hand side (*ps n* 5M) was by far the more common type, with falling F_0 in both stressed syllables. Part of the fall may be explained by the open glottis in the preceding consonants ([p] and [s] respectively), but not all, and the falling F_0 contours in these utterances also gave an impression of falling pitch, that is, utterance *ps n* 5M has two falling pitch contours while utterance *ps n* 2F has level pitch on the first syllable. In most cases there were no obvious rhythmical differences between the two types above; falling F_0 contours in the first syllable were not followed by pauses or noticeable lengthening of the vowel or final consonant. A few speakers (especially 4M) did have such realisations with apparent phrase boundaries after the first stressed word in the declarative sentences (never in the interrogatives *pdp* and *dsi*), although this was more common in the longer utterances.

A similar difference in the realisation of the first stressed syllable can be seen in the two typical realisations of sentence *bsa n* in Figure 2.7. In these two versions of *bsa n* one has rising F_0 in the first syllable and the other has level F_0. It is difficult to make direct comparisons between these examples and the previous ones of sentence *ps* because of the different segmental conditions, but sentence *bsa n* 2F illustrates the rising F_0 on the first (fully) stressed syllable which has often been noted in other studies. (In non-instrumental studies the rising *pitch* has been observed.) Although the rising initial F_0 was common in sentence *bsa*, some speakers had level F_0 in this position, as in *bsa n* 4M, or slightly falling F_0 (speaker 6M). Some speakers alternate between level, slightly falling and slightly rising F_0, with very little difference in the perception of pitch. Such (perceptually insignificant) alternations have sometimes been averaged, resulting in level F_0 contours.

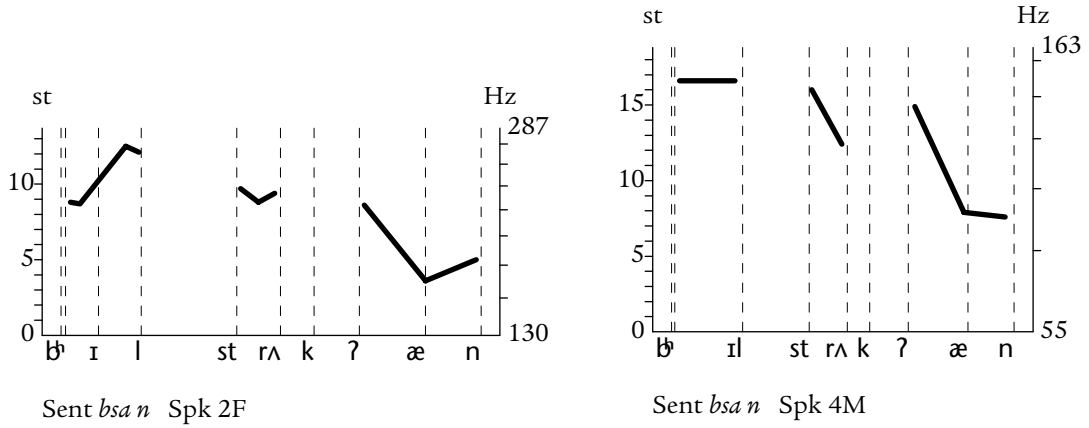


Figure 2.7. Sentence *bsa n* by speakers 2F and 4M, showing a difference between rising F_0 and level F_0 on the first stressed syllable.

The overall F_0 contour of the sentences is falling, each stressed syllable being 2–3 semitones lower than the previous one. The same tendency is found in sentence *jkft n* by speakers 5M and 6M in Figure 2.8.

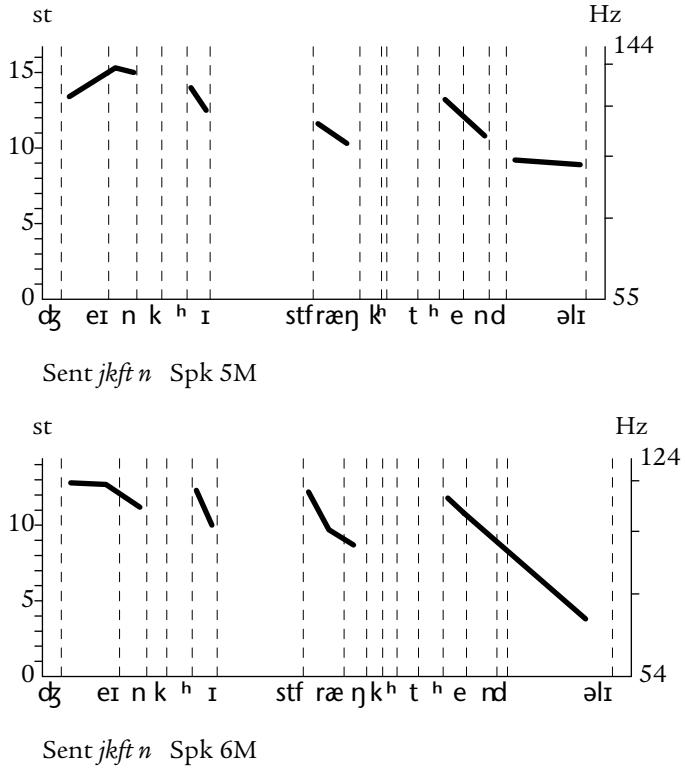


Figure 2.8. Sentence *jkft n* by speakers 5M and 6M. Here too there is a slight difference in the initial F_0 movement on the first (stressed) syllable. The rising F_0 pattern was particularly common in the longer utterances but always subject to inter-speaker variation.

The successive F_0 'step-down' is generally smaller in the two versions of sentence *jkft*, but this is partly speaker dependent. The corresponding sentences by speakers 2F and 4M show larger differences between the stressed syllables due to the larger F_0 ranges employed by these speakers. The final stressed syllable (the 'nucleus') appears

to be less downstepped than the preceding stressed syllables, even if we take the influence of the aspirated initial [tʰ] into account. The preceding stressed syllable onset is also a consonant produced with an open glottis ([f]). See also the comments on sentence *sepc n* below.

2.5.2.2 Neutral utterances without stress clash

The only systematic differences between the sentences presented here and those in the previous section is the presence of one or two unstressed syllables between each stressed syllable. The main point here is therefore how these unstressed syllables are placed in the F_0 contour and whether they seem to influence the overall F_0 contour. The sentence with two lexical items (*pc*) is represented in Figure 2.9 by the averaged traces from the same two speakers as for sentence *ps*.

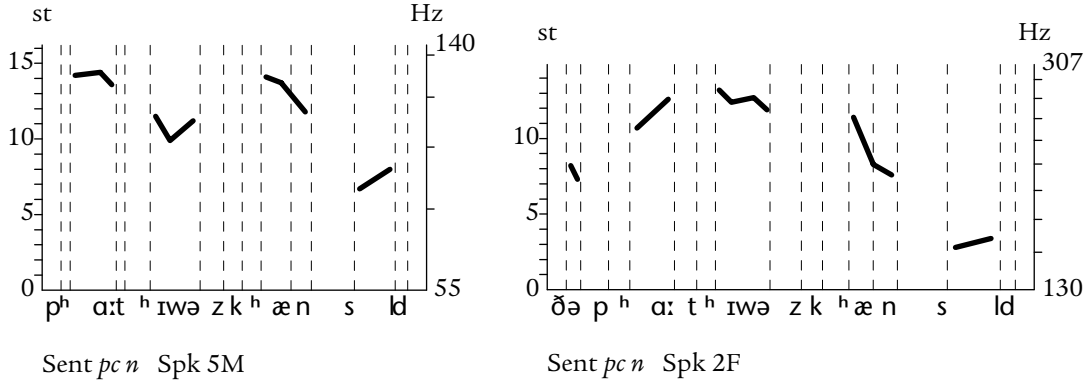


Figure 2.9. Sentence *pc n* by speakers 5M and 2F. Compare the F_0 contours with those of the corresponding sentence with stress class in Figure 2.6.

Sentence *pc n* by speaker 2F has rising F_0 on the first stressed syllable and F_0 stays high (more or less) throughout the following unstressed syllables before it falls on the second stressed syllable. This contour of a rise followed by a plateau and then a fall has been noted for several languages, including English, and was named the 'hat pattern' by Cohen and 't Hart (1967). This pattern does seem similar to the sustained F_0 on the one syllable 'Paul' in sentence *ps n* – at least they share the plateau plus fall, but the initial rise is absent in *ps n*. Speaker 5M uses the same pattern in sentence *pc n*, with unstressed syllables, as in *ps n* with stress clash, namely falling F_0 associated with both stressed words, only here distributed over more than one syllable. Both sentences (that is, *ps n* in Figure 2.6 (left) and *pc n* in Figure 2.9 (left) by speaker 5M) have what might be considered two falling accents in the traditional British system or two H(igh) – L(ow) accents in the Pierrehumbert system (probably H*+L H*+L), but the association with segmental structure differs.

The two versions of *css n* in Figure 2.10 are fairly representative of all speakers in having gradually descending F_0 throughout the sentence and a steeper fall on 'soup'.

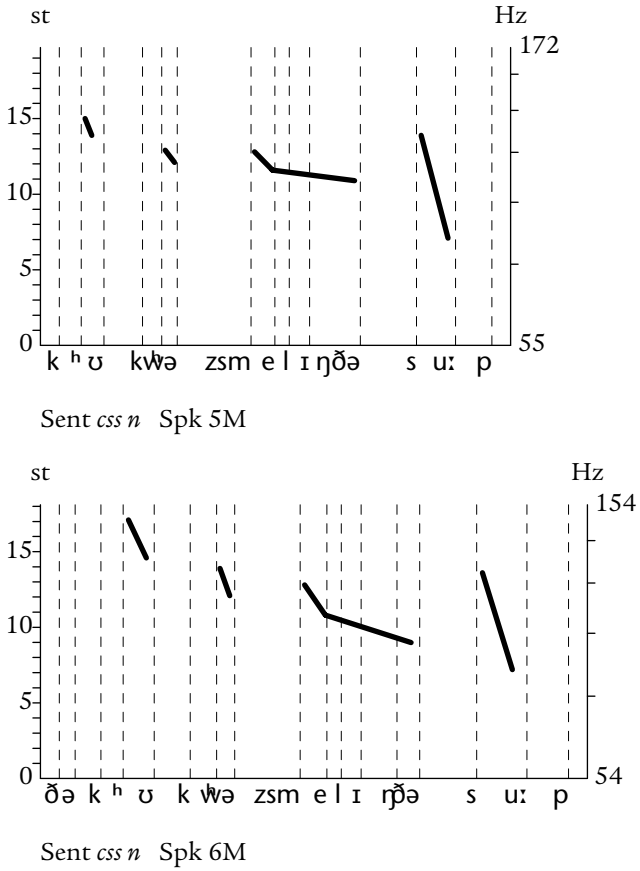


Figure 2.10. Sentence *css n* by speakers 5M and 6M. No reliable F_0 measurements (and for speaker 5M also segment durations) for the word 'The' could be obtained.

The first stressed syllable has falling F_0 for both the above speakers, but speaker 2F had rising F_0 on this syllable. Notice again the downstep from first to second stressed syllable but a smaller or no downstep in F_0 from the second to the final stressed syllable. The same tendency appears in some speakers' production of the sentence with four lexical items, as can be seen in Figure 2.11 (speaker 5M).

F_0 on the second and third stressed syllables in *sepc n* 5M is 3 and 2 semitones lower than on the preceding stressed syllable respectively, while the final stressed syllable – the nucleus – has slightly higher F_0 than the preceding stress, under similar segmental conditions. Speaker 2F has a different pattern with what appears to be more equal downsteps throughout the sentence.

The F_0 relation between stressed syllables and the following unstressed syllables seems to be subject to speaker variation as well as sentence context. For speaker 2F the unstressed syllables are generally, but not always, placed on the general downtrend line which also carries the stressed syllables, but for others, especially 4M, F_0 is typically lower on the unstressed syllables, which results in steeper slopes of F_0 from stressed to unstressed syllables overlaid on a less steeply declining downtrend line.

I mentioned earlier the rising F_0 on the first stressed syllable of an utterance. It is interesting that this is found more often in longer utterances than in shorter utterances with two stressed syllables, and vice versa that the shorter utterances are often, if by no means always, produced with two falling F_0 patterns. Such clear F_0

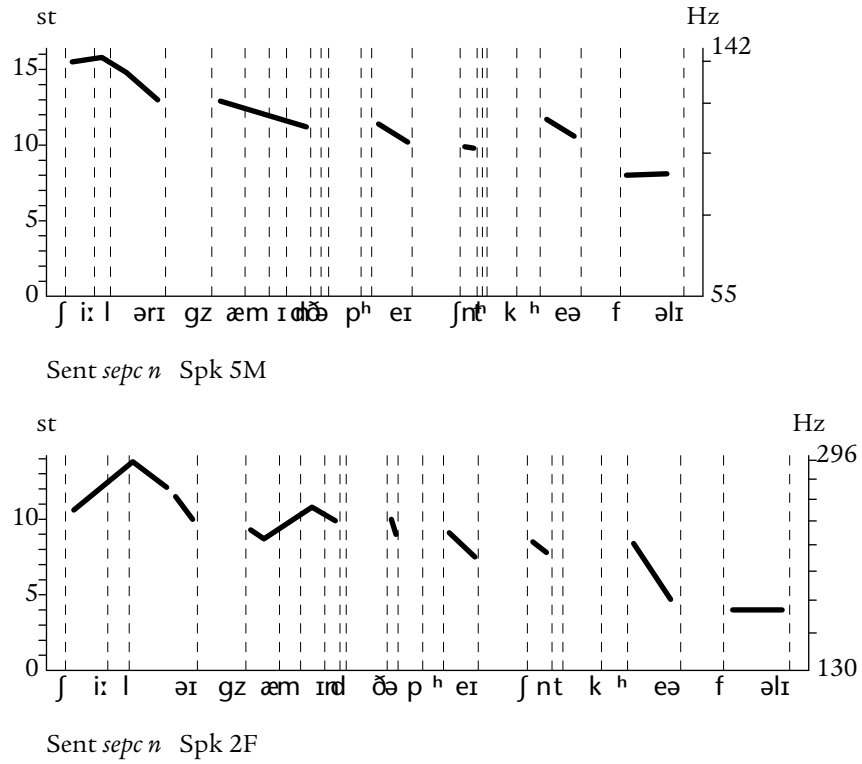


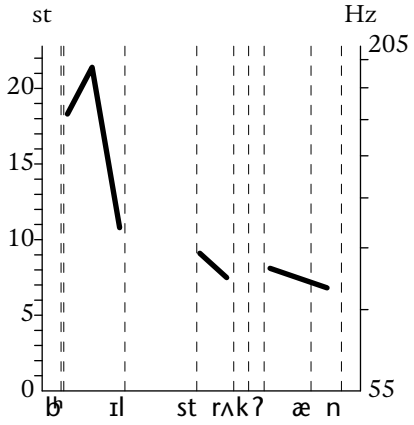
Figure 2.11. Sentence *sepc n* by speakers 5M and 2F.

movements are normally associated with nuclei in the British tradition, which would indicate that short sentences such as 'Paul sings' and 'The party was cancelled' are typically produced with two nuclei, that is, as two phrases. But why would such utterances contain two phrases? It seems ill motivated. And these utterances typically do not sound bi-phrasal. This difference in F_0 movements near the beginning of the utterance may be indicative of different planning strategies. The F_0 (or pitch) level of the first stressed syllable can serve as a an anchor point for the interpretation of pitch and prominence levels later in the utterance (Gussenhoven, Repp *et al.* 1997), and variations in initial F_0 peak height may possibly act as a signal to utterance length (in time or number of stressed syllables). Such planning information may be more important in longer utterances than in short ones. It is also possible that the rising F_0 in longer utterances is caused by a greater tolerance for 'undershoot' or delay which cannot be afforded in very short utterances.

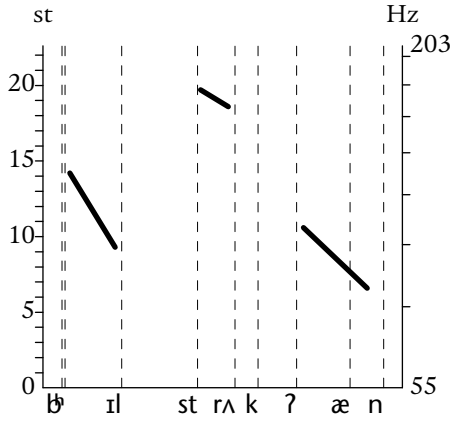
2.5.2.3 Marked information structure

The influence of narrow focus on the F_0 contour of an utterance seems quite consistent across speakers and sentence length, so the basic principles can be illustrated by the averaged traces for one speaker which makes direct comparison easier. Figure 2.12 shows the stress clash sentence *bsa* spoken by speaker 4M with focus on each of the three lexical items respectively, and Figure 2.13 shows the similar results for the

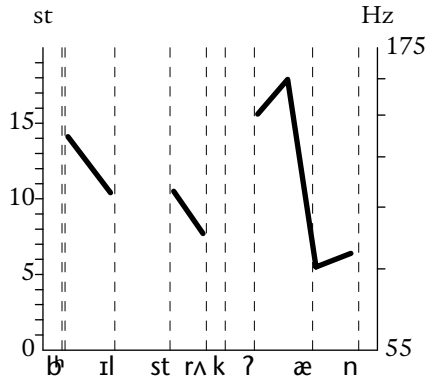
sentence *sepc* with four lexical items and intervening unstressed syllables, spoken by speaker 1F.



Sent *bsa f1* Spk 4M



Sent *bsa f2* Spk 4M



Sent *bsa f3* Spk 4M

Figure 2.12. Sentence *bsa* produced by speaker 4M with focus on the first (*f1*), second (*f2*) and third (*f3*) lexical item respectively. The corresponding neutral version of *bsa* is displayed in Figure 2.7.

It is clear from Figure 2.12 that the stressed syllable of the item which has been emphasised for (semantic) focus has much higher F_0 than any of the other syllables in the sentence or the corresponding syllable in the neutral version; F_0 is boosted upwards on the focal accent – in this case by 3 – 5 semitones compared with the neutral version. The non-focal items have lower F_0 than the same items in the neutral version, for this speaker similarly 3 – 5 semitones. Speaker 4M has the largest F_0 range of all six speakers, so the absolute differences (in semitones) are smaller for the other speakers, but the general pattern is the same. The non-focal items are produced with some F_0 movements by this speaker, even in post-focal position in version *f2*, although the movements are somewhat compressed in range in *f1*. The relation between the pre-focal items in *f3* and the pre- and post-focal items in *f2* is similar to that in the neutral version (only lower in the F_0 range) in being placed on a descending slope, while the two post-focal items in *f1* are (almost) equally low. This pattern becomes more apparent from the traces of sentence *sepc f(1-4)* by speaker 1F

in Figure 2.13.

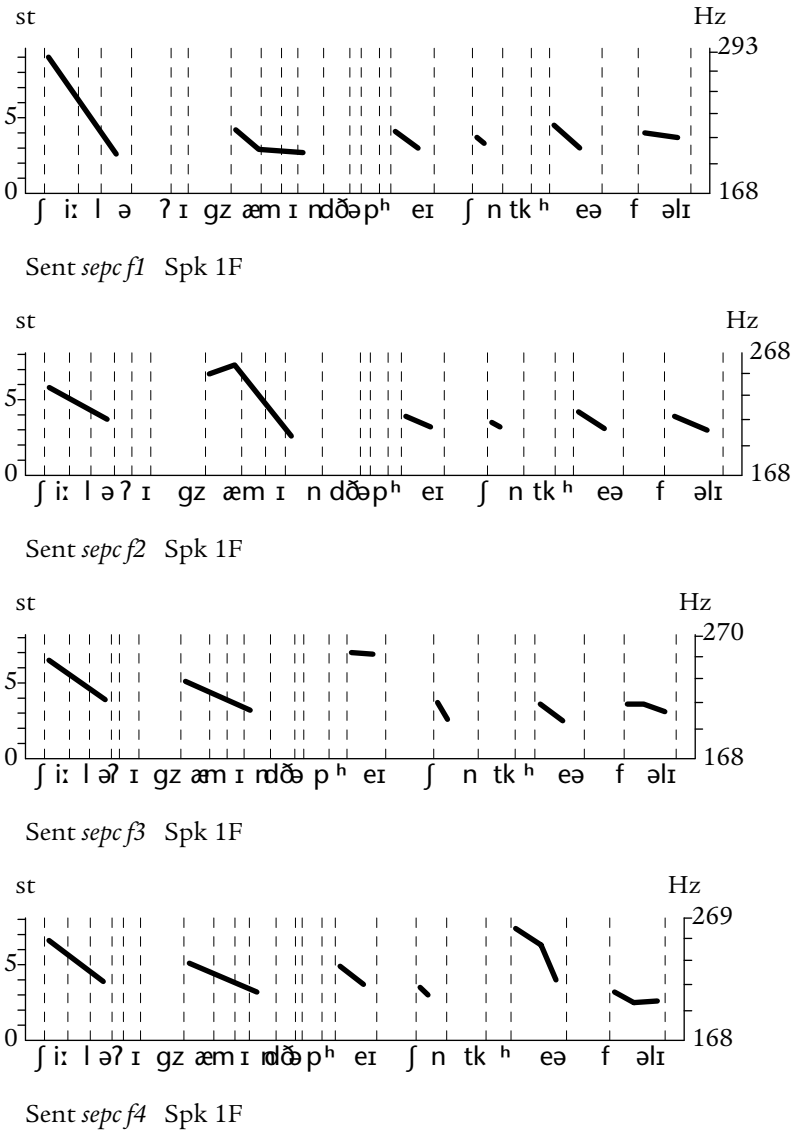


Figure 2.13. Sentence *sepc* produced by speaker 1F with focus on each of the four lexical items in the sentence, respectively.

Speaker 1F produces all four versions of *sepc f* with a relatively large F_0 fall on or around the focal accent, and F_0 then remains at this low level to the end of the sentence. Although some of the crucial data points are missing in these examples it seems that, in general, the F_0 fall is distributed over the stressed syllable and the following unstressed syllable (the first post-tonic syllable). The F_0 traces from the other speakers confirm that F_0 does not normally continue to fall after the first post-tonic; in some cases it even rises slightly.

As in the previous example (and in the material in general) F_0 on the pre-focal stressed syllables forms a descending line, and the relation between the stressed (non-focal) syllables and the following post-tonic syllable is the same as in the

neutral versions: for some speakers (such as 1F) a slightly steeper fall than between successive stressed syllables. The rising F_0 which was found on the first stressed syllable of some neutral sentences (especially *jkft* and *sepc*) is also present in some of the corresponding focused sentences, when the focal accent is near the end of the sentence, which means that this F_0 pattern is not completely lost through pre-focal F_0 reduction. This means that immediately adjacent items are influenced the most by a following focus.

2.5.2.4 Interrogative sentences

Two interrogative sentences were included in the material with the purpose of eliciting rising or falling-rising pitch patterns in addition to the expected falling patterns on the declaratives. Only three and four speakers were consistent enough to allow pooling of their productions of the neutral versions of the two sentences *pdp* and *dsi* respectively.[3] The three speaker versions of *pdp n* all had different F_0 patterns and would probably be classified differently in both the traditional British system and in autosegmental-metrical frameworks such as Pierrehumbert (1980) or ToBI (see Section 1.6). They are shown in Figure 2.14.

The three versions of *pdp n* vary with regard to (1) the steepness of the F_0 slope and (2) the F_0 movement on (or associated with) the nucleus. Speaker 4M has the steepest descending slope, but this is a general characteristic of this speaker due to his large overall range. Both he and speaker 2F have the same F_0 configuration in the first part (the 'head') of this sentence as in the declarative sentences described above, but the final stressed word is different. Speaker 2F has falling-rising F_0 on this word, but the resulting auditory impression is a low-rise (in the British tradition, Cruttenden 1997: 50), that is, the initial fall is not perceived (as a significant fall), rather the stressed syllable is perceived as low and the following unstressed syllable as high(er). The overall F_0 contour in the sentence corresponds to Armstrong and Ward's 'Tune II' (1931:19). Speaker 4M has falling F_0 on the final stressed syllable followed by a small rise. The rise is audible but very weak, and it is debatable whether it should be classified as a fall-rise or a simple fall. In case of the latter the F_0 contour in the sentence corresponds to Armstrong and Ward's 'Tune I', in line with most of the declarative sentences. The last version is different with regard to both the F_0 movement on the last stressed word and the slope of the preceding F_0 contour. The contour is less steeply declining than in the declarative sentences by the same speaker, and the rising F_0 on the nucleus is thus likely to be interpreted as a high-rise.

[3] Speaker 3F was also consistent, but her production of neutral utterances deviated so much from the norm that they have been excluded from most analyses. See Section 4.1 for a more detailed explanation of this.

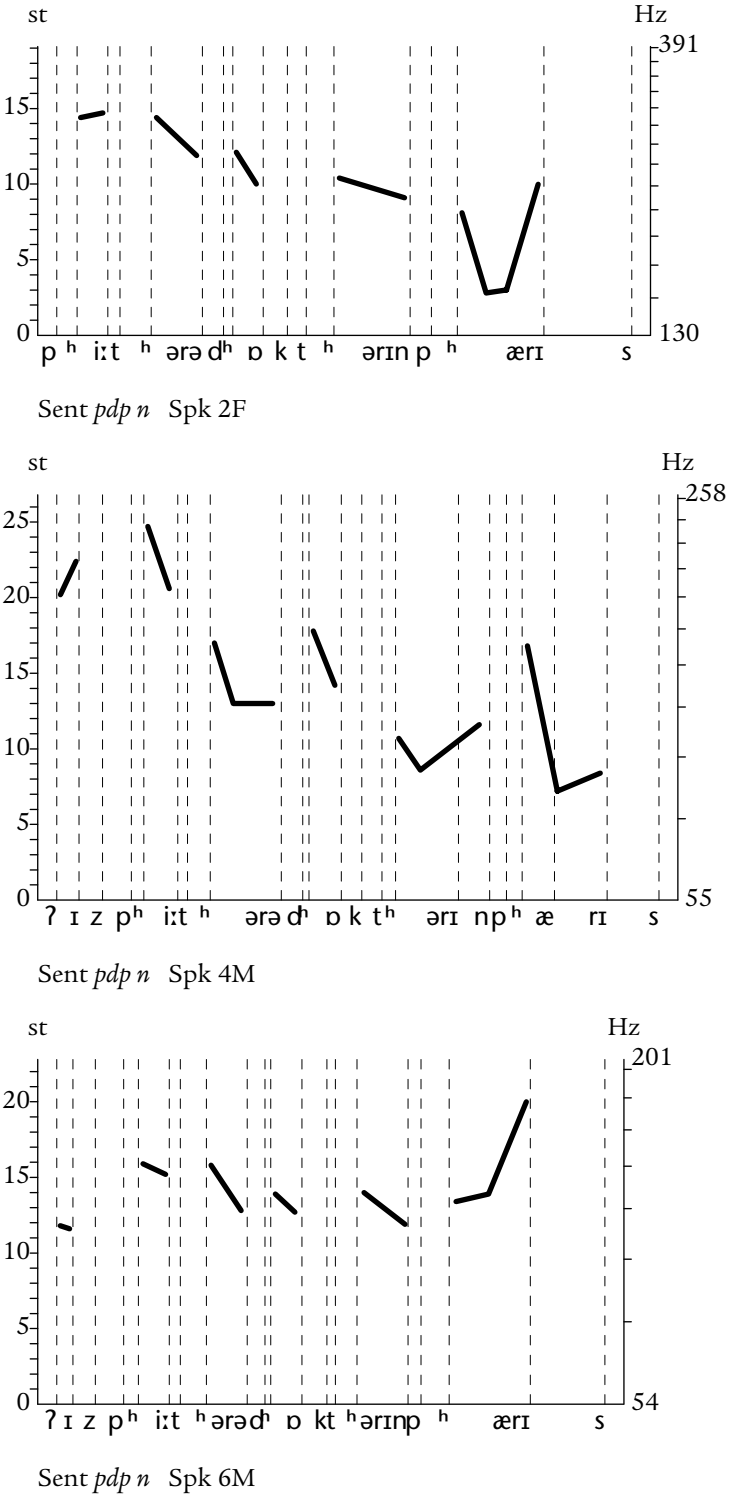
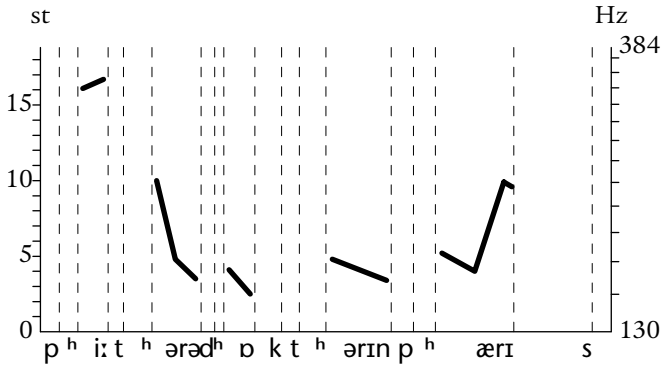
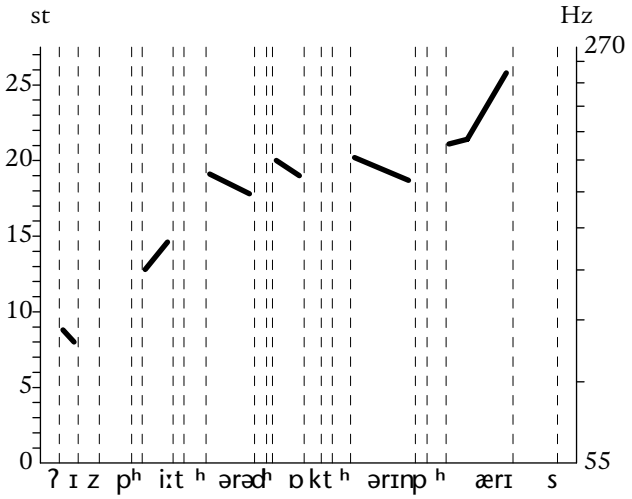


Figure 2.14. Neutral versions of the sentence *pdp* by three speakers. Perceptually they have low rising (2F), falling(-rising) (4M) and high rising (6M) pitch, respectively.

The corresponding versions with focus on the first item also showed some variation, but two very distinct types are shown in Figure 2.15.



Sent *pdp f1* Spk 2F



Sent *pdp f1* Spk 4M

Figure 2.15. Two versions of sentence *pdp* with focus on the first word. The one by speaker 2F has a 'fall-rise' distributed over the whole sentence, and the other, by speaker 4M has a 'high-rise' (following the terminology in Cruttenden 1997).

Sentence *pdp f1* 2F has very high F_0 on the stressed syllable of the focused word followed by low F_0 on the post-tonic syllable – a configuration which is perceived as a fall. F_0 then stays low until it rises near the end of the stressed syllable in the final word. This pattern is normally called a fall-rise in the British tradition, only distributed over the entire sentence.

In sentence *pdp f1* 4M the focused word has an associated F_0 rise from the stressed syllable to the following unstressed syllable. F_0 continues on a (relative shallow) incline until it rises sharply on the final stressed word. This pattern of (an early) rise plus level plus final rise is also well noted in the literature. The two above patterns can be seen in Pierrehumbert (1980), figures 1.5C (p. 262) and 1.6C (p. 264, respectively, where they are analysed as 'H* L- H%' (fall-rise) and 'L* H¯ H%'.
The realisations of sentence *dsi* follows the same patterns as *pdp*, only spread over a longer stretch of five lexical (and therefore stressed) items.

2.5.3 Duration and stress

As was mentioned in Chapter 1 duration has been shown to be both a clear acoustic correlate of and a strong perceptual cue to stress (Fry 1955, Adams and Munro 1978, Nakatani and Aston 1978, Silipo and Greenberg 1999, Silipo and Greenberg 2000).

The relation between duration and stress is normally assumed to be more straightforward than between F_0 and stress. It is (now) generally acknowledged that stress can be signalled both by low (rising) F_0, as in Standard Copenhagen Danish or low-rise tones in English, or high (falling) F_0, and that very similar F_0 configurations can result in different perceptions of the location of stress, depending on relatively minor differences in the timing of the F_0 movements or the interplay with other acoustic parameters. Duration is normally treated as a fairly mono-dimensional parameter: longer duration results in the perception of greater prominence, although some interaction with both intensity and F_0 is generally recognised. This does not mean that the relation between duration and stress is simple, only that the relation between the acoustic parameter and the perceptual parameter is relatively straightforward. One of the central issues regarding duration and stress, and where languages have been shown to differ, is the relevant scope or domain of duration differences between stressed and unstressed syllables and words, or what is sometimes called 'accentual lengthening'. It has been suggested that in Dutch the relevant domain of accentual lengthening is the entire word, and that relative duration between syllables within the word is constant (Sluijter 1995: 37), while Turk and Sawusch (1997) have suggested that the relevant domain in (American) English is the stressed (accented) syllable and the following unstressed syllable(s), while word-initial syllables are not affect by accentuation. Another difference between Dutch and English is the interaction between accentual lengthening and final lengthening. Cambier-Langeveld (1999) found a strong interaction effect for Dutch, so that the duration difference between accented and unaccented words in final position was very small, while the lengthening effect of accentuation was as strong in final position as elsewhere for English. This is interesting considering the findings of Nakatani and Aston (1978) that duration was not used as a cue to stress in final position in their study (see Section 1.2).

Below a brief account is given of the influence of stress and focal accent on duration in the present study. As with the description of F_0 above this is not a full-fledged and complete analysis. Many of the features which can effect the outcome of an analysis (such as the issues of the relevant domain(s) just mentioned) have been left for future study, but the general observations about the variations in duration will serve as a background for the perceptual studies later in the thesis.

Only the (lexically) stressed syllable of the lexical items in the sentences is considered, since the lexical words are the ones which occur in different focus conditions in the material. Informal inspection of the data suggested that the durational differences were often not clearly present in the vowel segment alone, but were spread over the vowel and (at least) any following sonorant(s). Therefore, and because the boundaries between vowels and following sonorants often cannot be determined with any accuracy or consistency, the duration measurements are based on the vowel plus (any) following sonorant consonant. There was one exception; the word 'Ann' in *bsa* is utterance final, and the exact endpoint was, as a result, often indeterminate. Instead, only the vowel **[æ]** was measured. The comparisons of duration are only

performed on the 'same', or corresponding, segments, that is, the duration of the vowel plus lateral **[ɔːl]** in the neutral version of 'Paul sings' is only compared with the same segment(s) in the same sentence in different focus conditions (neutral versus focused versus non-focal in focus sentence). Table 2.1 contains the mean ratios of duration in focused sentences (whether in focal or non-focal position) to duration in neutral sentences. This provides information about the effect of foregrounding and backgrounding in different positions in a sentence. The corresponding ratios for each individual speaker and the actual durations (in milliseconds) from which they were calculated are found in Appendix B, Section B.4.

Segment duration ratios – all speakers (mean)						
Sent	*Lex*	*Segment*	*f1/n*	*f2/n*	*f3/n*	*f4/n*
ps	1	ɔːl	**1.24**	0.80		
ps	2	ɪŋ	0.90	**1.15**		
pc	1	ɑː	**1.11**	0.89		
pc	2	æn	1.05	**1.06**		
bsa	1	ɪl	**1.17**	0.80	0.74	
bsa	2	rʌ	0.98	**1.09**	0.93	
bsa	3	æ	0.88	0.90	**1.09**	
css	1	ʊ	**1.12**	0.80	0.88	
css	2	e	1.02	**1.20**	0.95	
css	3	uː	0.93	0.94	**1.07**	
jkft	1	eɪn	**1.22**	0.83	0.76	0.80
jkft	2	ɪ	1.01	**1.08**	0.92	0.87
jkft	3	ræŋ	0.92	0.89	**1.12**	0.88
jkft	4	en	1.00	1.03	1.00	**1.08**
sepc	1	iː	**1.13**	0.84	0.84	0.83
sepc	2	æ	0.96	**1.30**	0.95	0.97
sepc	3	eɪ	1.00	1.03	**0.99**	0.95
sepc	4	eə	0.93	0.94	0.94	**1.06**
pdp	1	iː	**0.97**			
pdp	2	ɒ	1.00			
pdp	3	ærɪ	0.93			
Segment duration ratios of duration in focused context (*f1-4*) to duration in neutral context (*n*)						

Table 2.1. Segment duration ratios relative to mean duration in 'neutral, context-free' sentences. The duration of a segment is divided by the duration of the same segment in the neutral version. The figures are mean values of the ratios for the six individual speakers. Segment duration ratios in words which have been emphasised for semantic or focal contrast are printed in bold-face type.

Several trends appear from the duration ratios in Table 2.1. First, as expected the stressed syllables (vowel nucleus or rhyme) are generally lengthened when the word is focused compared with a stressed, neutral production of the same segment. The order of the increase varies, and there are segments which do not undergo lengthening, for example [iː] in 'Peter' (sentence *pdp*) or [eɪ] in 'patient' (sentence *sepc f3*). In most cases the segments are lengthened 5–15%. This lengthening seems to occur in all positions in the sentences with no clear differences in the magnitude of the lengthening.

Second, the non-focal segments in sentences with a specific focus are normally shorter than in neutral utterances. This reduction is often as large as or even larger than the lengthening of the segments with focal stress/accent. There appears to be a difference in the magnitude of this effect depending on the position of the item relative to the emphasised word: pre-focal segments are generally reduced more than post-focal segments. Note that in Table 2.1 position within a sentence is indicated from top to bottom. On the vertical axis pre-focal segments come before focal and then post-focal segments. Horizontally, that is, within one line, post-focal items *precede* the focal item, which is then followed by pre-focal items. A simple calculation of mean values shows that pre-focal segments (stressed syllable rhyme of pre-focal lexical items) reduce to 86% of their duration in neutral sentences, while post-focal segments only reduce to 96% of their duration in neutral sentences. This is the opposite pattern of what we saw earlier in the section about F_0, where it was clear that the F_0 movements were much more restricted in post-focal position than in pre-focal position. This may be an indication that these two parameters are employed differently in pre- and post-focal position, not only as acoustic correlates but also as perceptual cues. Temporal organisation, whether regarded purely as prominence through segment durations or as the somewhat intangible concept of speech rhythm, may be more significant pre-focally, while variations in pitch are more important in post-focal position.

CHAPTER 3

Validating prominence ratings

3.1 Introduction

The evaluation process described in the previous chapter provides a good indication of the 'sameness' (or not) of the individual utterances. When it comes to the specific judgements of, for example, the degree of stress on a particular word or word position, such as post-focally, it is less obvious that the judgement of one rater, such as the investigator in this study, will accurately reflect the opinion of the 'language community' in general. This is true whether the language community in question is native speakers of English in general, speakers of Southern Standard British English, professional (English) language teachers in Denmark, phoneticians, or other groups which are relevant for the present work. In fact, although a number of studies (Silverman *et al.* 1992, Heldner 2001a) have shown good agreement between raters when it comes to stress assignment, it was still considered desirable to demonstrate such agreement for the present material. An evaluation of some of the relevant parameters was therefore obtained from a number of listeners from the relevant language communities by means of listening, or perception, experiments (especially Test 1), supplemented by some experiments which were prompted by the results of the first ones. The experiments, referred to below as Tests 1–4, focused on the two main parameters under investigation here: 1) *prominence*, in the sense of perceptual salience, that is, how much a syllable stands out by virtue of its physical properties (in a wide sense including its position in the utterance), and 2) *information structure*, that is, whether listeners are able to decode the intended focus information.

The listening experiments attempt to answer several questions which pertain to the perception of stress and phrasing in the utterances in general. First, is there sufficient inter-listener agreement on stress levels and information structure to justify general statements about the physical manifestation of these phenomena? This of course was an absolute requirement for the project, both in relation to the analysis of perceived prominence and to any analysis of the acoustic correlates of the perceived prominence. Second, if such an agreement does exist, can the evaluation of (any) one rater be said to be representative of the 'general opinion' of native English speakers or of other relevant language groups? In other words, is it reasonable to base an analysis of prominence on the perception of one rater only? As mentioned above, Test 1 was designed specifically to address this issue. And third, what are the systematic differences in perception of stress levels or focus structure, if any, in relation to for

example position in the utterance – first, second, or third stressed word, or pre- or post-focal position? Each individual analyst may have more or less strong intuitions about this, but a controlled experiment with a group of listeners can reveal if these intuitions can be generalised.

3.2 Material

The same stimuli are used in all the (sub-)experiments, although only a subset of utterances was selected for Test 4. The material is rather large; one token was extracted from each group of what had been deemed to be 'perceptually equivalent' utterances, from each of the six speakers, giving a total of 183 utterances. The selection of individual tokens was not based on their auditory impression – they were simply the first repetition in the series of perceptually equivalent utterances. One of the utterances had to be replaced by another repetition because of some distracting microphone noise.

The general impression of overall loudness varied somewhat from speaker to speaker because of slight differences in recording level and, more importantly, differences in speaker reading style and overall loudness. This was felt to be a possible source of error, or at the very least a distracting factor, in an experiment which aims to investigate the perception of prominence, so the sound files were modified in the following manner.

The maximum peak intensity of each sound file, containing one utterance, was extracted from the file header,[1] and the average peak intensity across files was calculated for each speaker. The speaker with the highest average peak intensity was used as a reference, and a correction factor was calculated for each of the other speakers according to the following simple formula:

$$CF_x = \frac{Intensity_{max}}{Intensity_x}$$

where
CF_x = correction factor for speaker $_x$
$Intensity_{max}$ = average peak intensity for 'loudest' speaker
$Intensity_x$ = average peak intensity for speaker $_x$

The sound files were then converted from ESPS format to Microsoft® RIFF WAVE audio (or .wav) format using the tool CopyAudio from the AFsp package[2] while applying the calculated gain factor (by multiplication). An impressionistic auditory check of the output sound files revealed no obvious differences in loudness.

[1] ESPS® FEA_SD data files.

[2] See http://www.tsp.ece.mcgill.ca/Docs/Software/AFsp/AFsp.html for a description.

3.3 Tests 1–3 – perception of prominence

Tests 1–3 can all be considered sub-experiments of the same overall experiment with regard to method and material; the presented stimuli are the same: the exact same utterances in the exact same order. However, they differ with regard to the instructions given to the raters and/or with regard to the linguistic background of the raters. Since these differences are also under investigation and because the tests were carried out as separate experiments where the results of one test led to the formulation of the next they are presented here as separate experiments which are obviously closely linked.

The 183 utterances were arranged in a randomised list, split up into four sections and presented to the raters on a webpage. The actual performance was left almost entirely to the raters' discretion: they could 'click' on a sound file to have it played back according to the regular setup of their respective browsers; they could adjust the sound level and listen to the utterances as many times as they wished, and they were allowed to complete the test at their own pace – in larger or smaller instalments. The only requirements were that they listen to the utterances in the order in which they occurred on the webpage, and that they indicate whether they used loudspeakers or headphones. The ratings were recorded on printed 'answer sheets' using a fairly conventional notation style (see the individual tests for a precise description).

3.4 Test 1 – Danish listeners

3.4.1 Subjects

Running this test with non-native, here Danish, speakers of English was not originally by design only but also by necessity. Finding native speakers of British English who are both able and willing to do a fairly long and difficult listening experiment is difficult when one is not based at a linguistics department at an English University. However, there are also good non-practical motivations for using Danish raters. One of the objectives of this project is to establish the differences between the realisation of prominence relations in Danish and in English, or at least to establish a foundation for investigating these differences. It is therefore highly relevant to know how Danish listeners perceive these differences, irrespective of how they are manifested physically, or acoustically. Furthermore, during the search for suitable subjects it became clear that my candidates belonged to different groups in terms of linguistic background and training, and I suspected that this could lead to systematic differences in the way they would respond to the stimuli. Linguistic training and background was therefore included as a parameter in the listening experiment.

There are three groups of listeners in the test:

(1) phoneticians trained and working in general phonetics, primarily with Danish as their target language
(2) phoneticians from English (language) departments in Copenhagen (University and Business School), trained and working in English phonetics

(3) advanced phonetics students from the Linguistics Department at the University of Copenhagen.

There were three listeners in each group, and in addition I performed the test myself, giving a total of ten listeners, or raters. All the raters in groups 1) and 2) have at least 15–20 years of professional experience in phonetics, including prosody, and the students in group 3) had all had some experience with marking and transcribing Danish, but not English, prosody. All the raters are competent speakers of English (as a second language), ranging from intermediate (or advanced) learner level to native-like proficiency. My own background is mixed; my training is in English phonetics, but I have also worked with Danish phonetics in research and teaching and have about ten years of experience in phonetics.

3.4.2 Purpose of the listening test

The purpose of this experiment is threefold:

(1) Does the type and perhaps amount of linguistic training influence raters' perception of prominence in a systematic way? The presence of such an effect would be an indication that we may adapt our perception to the theoretical framework to which we subscribe, and not, as should ideally be the case, only the other way around.

(2) What are the variations in perceived prominence levels in relation to the parameters which were included in the reading material, such as number of stressed words/syllables in a sentence, position of the stressed word or of a focused element etc. Having ten raters evaluate the material makes it possible to make averages of the ratings (supposing there is sufficient inter-rater agreement), and these averages may reveal far more subtle variations in prominence level perception than would be possible with any one listener. These averages may give rise to the formulation of certain specific hypotheses about variations in prominence levels, but it is already possible to formulate some questions based on the informal evaluation (See Section 2.3) and the acoustic characterisation (especially of F_0) in Section 2.5. They are:

    a) The degree of reduction of a stressed word/syllable is related to its proximity to a focused element.
    b) The first stressed syllable (the onset) and the last stressed syllable (the default nucleus) in neutral, context-free utterances will be more prominent than intermediate stressed syllables.
    c) Stress reduction is more pronounced in post-focal position than in pre-focal position.

(3) Do the ratings of any one rater, in this case particularly the author, reflect the prominence ratings of the group as a whole in a satisfactory manner? This

might be relevant if a larger part of the utterance material, which has not been assessed by multiple judges, is to be analysed with regard to the connection between perceived prominence and acoustic properties.

Originally this experiment was designed to address questions 1) and 3) in particular, using just one set of raters, but the preliminary results of the attempt to answer question 2) led to a more detailed analysis of this issue, presented in Chapter 4, and the addition of further tests, presented in Chapters 5 and 6.

3.4.3 Instructions to the raters

The instructions to this test were in Danish and as such cannot be compared directly to the instructions given in the experiments with native English speakers. It was a matter of great concern to find formulations that would lead the raters to mark prominence in a manner which is dictated as little as possible by whichever theoretical framework they are familiar with. This is a non-trivial issue, especially when the raters come from very different linguistic traditions and have different language backgrounds, namely Danish versus English, and is also discussed in Section 5.3, in connection with Test 2. The choice of formulation for the Danish raters was relatively easy. There is a very well established terminology which I believe most native speakers of Danish readily associate with what I wanted them to mark, namely linguistically relevant prominence. The Danish equivalent of the word 'stress' is *tryk*, and it is used to cover both lexical stress and actual utterance level prominence. Perhaps because of the absence in Danish of what some linguists would call different 'pitch accent types', with different tonal configurations (e.g. Ladd 1996: 83) or of an obligatory nucleus as described in the British tradition (e.g. Cruttenden 1997) there is no distinction between 'stress' and 'accent' in Danish, and the latter word is absent from both everyday and academic usage. If a word has *tryk* (stress) it is perceptually prominent. This explanation reflects my own intuition about stress in Danish and the intuition of those with whom I have discussed it. It is, I believe, uncontroversial.

It is a matter of some controversy, though, just how many levels are needed in a phonological description – just as is the case for English – but everyday usage indicates that *tryk* (stress) is normally felt to be continuously variable, that is, a matter of more or less, which is in good agreement with a phonetic definition of stress (as prominence or degree of emphasis). In this test I asked the raters to mark three degrees of stress above completely unstressed. This decision was based in part on the four-level hierarchy of stress and accent which is assumed in some descriptions of British English (Gimson 1989, Cruttenden 1997), and with which I wanted to compare my results, and in part on an intuitive feeling that (at least) this many levels are needed for an adequate representation of the prominence relations in some of the utterances. The Danish instructions are available online (http://www.cphling.dk/pers/chrjen/stress/stress.html), and can also be found in Appendix A, Section A.3.

Here is a translation of the important parts (the original layout is maintained where possible):

Prominence:

Three levels of prominence are distinguished:
- Strong stress, indicated by two vertical strokes, e.g. "Peter did it.
- Normal stress, indicated by one vertical stroke, e.g. 'Peter.
- Weaker stress, indicated by one lowered stroke, e.g. ˌPeter "didn't do it.

Weak stress (= no stress) is not indicated.

Prosodic boundaries:

Two levels of prosodic boundaries are distinguished:
- Strong prosodic boundary, marked by //.
- Weaker prosodic boundary, marked by /.

Since the material consists largely of very short utterances it could not be expected that the raters would perceive many prosodic boundaries, and this prediction was borne out. Therefore the boundary responses are not analysed systematically but are used, where applicable, to cast light on differences in rater responses which are related to issues in which prominence and phrasing are strongly interconnected, such as marking of information/focus structure.

3.4.4 Listener feedback on the test

As mentioned earlier the test was fairly long. Although I was very familiar with the material and confident about the task it took me just under one hour to complete it. Many of the raters in the experiment needed more time – from one to two hours – and reported that they found the task quite difficult. One of the raters (*r* 8) reported that he was not able to make a proper distinction between 'normal stress' and 'weaker stress', and that he had stopped trying to make that distinction after the first few utterances. This is completely within the guidelines of the experiment and will simply be taken to indicate that the rater did not perceive weaker stress on (almost) any syllable. As will appear from the analysis of the data this distinction was in general a very problematic one.

3.4.5 Data

In order to be able to do a quantitative evaluation of the data they were coded in the following manner:

Strong stress	= 3
Normal stress	= 2
Weaker stress	= 1
No stress	= 0

Although weak, or no, stress was not marked explicitly by the raters, absence of a stress mark was assumed to signal an implicit marking of no stress which could thus be assigned a value. It is possible that responses would have been different if the raters had been asked to point out the unstressed words/syllables explicitly, but these (possible) differences seem to be of theoretical rather than practical importance, and it would have complicated an already difficult task unnecessarily.

Because of the size of the test it is not possible to gather much information by simply inspecting the raw data. The test material contains 907 words which have all been assigned a value by each of the ten raters, giving a total of 9070 data points, excluding the boundary markings. A large part of the evaluation process therefore consists in reducing and summarising the data to make the general trends appear more clearly. Table 3.1 shows a small section of the data file, with some of the information left out to make it easier to read:

	Raters									
Word	*r*1	*r*2	*r*3	*r*4	*r*5	*r*6	*r*7	*r*8	*r*9	*r*10
Bill	2	2	2	2	2	2	3	2	2	2
struck	0	0	1	1	1	0	1	0	0	1
Ann	2	3	2	2	2	3	2	2	0	2
Sheila	2	2	2	2	2	2	2	2	2	2
examined	3	3	3	3	3	3	3	3	3	3
the	0	0	0	0	0	0	0	0	0	0
patient	2	1	1	2	1	2	1	1	1	1
carefully	2	2	1	2	1	2	1	1	1	1

Table 3.1. Excerpt (edited) from the raw data file, showing listener responses for two utterances, 'Bill struck Ann' (*bsa*) and 'Sheila examined the patient carefully' (*sepc*).

The information which has been excluded from this sample was mainly used to identify each word, or token, according to speaker, intended focus and position of the token in the utterance. Organising the response data in such a manner, with separate lines for each word, implicitly makes the assumption that prominence assignment for one word is independent of the utterance in which the word occurs. This may not be appropriate for all types of analyses, resulting in unduly large degrees of freedom in some cases, but since the test instructions made no restrictions on the number or

sequence of each type of prominence label that is permitted in an utterance, they are in theory independent. A first inspection of the sample reveals that the responses are at least not completely random: for the three words 'Sheila, examined, the' there is total agreement between all raters, and the ratings for the other words differ by only one degree of stress.

3.4.6 Testing reliability and agreement

The reliability of rater responses can be tested in several different ways, depending on the kind of reliability or agreement required for the purpose. Evaluating the responses of a group (or several groups) of raters can be likened to the process of evaluating the accuracy of technical measuring instruments such as a speedometer or a sound level meter. In fact, when a group of listeners make subjective judgements about the *degree* of prominence in a speech sample they act as measuring instruments in the obvious absence of a mechanical instrument. Such a measuring instrument needs to meet various requirements before one can conclude that the extracted information is a true expression of general tendencies in the data material. The perceptual experiments in this chapter were in part inspired by similar experiments by Heldner (2001a), and I have used some of the same statistical techniques, as outlined in chapter 5 of T. Rietveld and Hout (1993) and the various articles and books mentioned there, although some modifications were made which will be explained below. Following Lawlis and Lu (1972) and others I will make a distinction between inter-rater *reliability* and *agreement*. To quote from T. Rietveld and Hout (1993), 'the concept of reliability is directly related to the extent to which measuring instruments covary, i.e. give relative values which are correlated' (p. 188). In other words, this is a measure of whether the listeners could perform the task in such a manner that the ratings from one listener could be predicted with some confidence from the ratings of another listener, even if the ratings are not identical in terms of absolute scores. The relevance for these experiments is that if the prominence judgements for the words of e.g. one utterance covary between several raters, one may assume that (1) the concept of prominence is at least to some extent uniform across the group of listeners, and (2) they perceived the same relative prominence. If we want to be able to make statements about the nature of the realisation of the different degrees of stress and/or accent we may require that the ratings not only covary, but also that they agree with regard to the absolute values, cf. the following definition: 'Agreement can be defined as the extent to which instruments return identical values' (T. Rietveld and Hout 1993: 188). If the raters do not agree about the absolute values in their judgements of prominence levels it may reflect differences in their understanding of what constitutes 'normal stress', or 'weaker stress' or one of the other labels or categories which they were asked to assign to each word, according to the instructions about the experiment, or it may reflect a lack of ability to make a systematic distinction between two such categories, in which case it will also show up in the test for reliability.

3.4.6.1 Reliability

Testing the reliability of the ratings involves decomposing them into a 'true score' and an 'error' component. The true score is the mean score for a very large (preferably infinite) number of raters, and the error is an expression of the deviation from this score. The measurement error can consist of several different variance components depending on the assumptions made about the data, and the reliability index is calculated as the ratio between the variance of the true scores and this variance plus the measurement error. Therefore, the larger the error the lower the reliability index.

This type of analysis requires, or assumes, data measured on an interval scale, and it may be debated whether this requirement is actually met here. Presumably the prominence ratings can be placed on a scale according to *degree* of stress, or prominence, but it is doubtful whether they are equidistant. Such a claim certainly cannot be sustained in its pure form. The distance in time from 1800 to 1900 is exactly twice the distance from 1900 to 1950, but we cannot claim that the difference between no stress (0) and normal stress (2) is twice as large as the difference between normal stress and extra strong stress (3). However, we may still use statistics which assume interval-level measurement for the analysis. Tinsley and Weiss (1975) state (reporting results from a paper by Baker, Hardyck and Petrinovich) that

> statistics which assume interval-level measurement can be applied to ordinal data without distorting the sampling distribution. Thus the [investigator] may appropriately use interval-level statistics on ratings that result from ordinal level measurement.

They do warn that

> The researcher should be aware, however, that such applications might result in measures of inter-rater reliability and agreement that distort the true relationship among raters, if the assumption of equal interval is grossly inappropriate. (pp. 360-61)

In other words, unless the analysis reveals that this is an untenable position – that the intervals clearly are not equal – the data can be treated as if they were measured on an interval scale.

The reliability coefficient which is normally used for this type of data is Cronbach's α (alpha), referred to in T. Rietveld and Hout (1993) as $R_{k(\mathrm{f})}$. The procedure is described in Winer *et al.* (1971: 1011-1015). The subscript k(f) indicates a group of raters k, who are considered to be a fixed factor (f), because they have not been randomly selected, but are specially chosen experts. The coefficient $R_{1(\mathrm{f})}$ expresses the typical reliability of a single rater, that is, the expected correlation between the ratings of one rater and those of another rater. The calculation of both coefficients is based on a two-way analysis of variance with one observation per cell.

3.4.6.2 Agreement

There are several tests or indices one can use to indicate the level of agreement among the raters, depending on the definition of agreement. One of the commonly used procedures (Silverman *et al.* 1992) is to make pairwise comparisons between all possible pairs of listeners in the group for each of the tested variables – here prominence scores for each word. Agreement is expressed as a quotient or percentage of pairs which agree in relation to the total number of pairs. For example: if we have three raters, [a, b, c], this results in three possible pairs, [a-b, a-c, b-c] (the order of the two raters is irrelevant). If rater [a] gave the score 1, and raters [b] and [c] the score 2, then only one pair – [b-c] – will agree, giving an agreement rating of 33%. Note that with only three pairs there are only three possible agreement scores, namely 0% (three different scores), 33% (two raters agree) and 100% (all three raters agree). With ten raters we get 45 possible pairs and a much finer gradation in agreement levels. Note, though, that when the number of raters exceeds the number of scale points (in my case: four) it is obviously not possible to have 0% agreement.

On the definition of agreement

It is possible to operate with stricter or more relaxed definitions of agreement. If the scale has many levels, or scale points, agreement can be defined as being +/– one scale point, or even more than that. However, with only four scale points such a relaxation of the requirements does not seem warranted, so in all the experiments described in this chapter, agreement will be defined as a perfect agreement in scores.

Exact agreement

A different index is often proposed as a strict and reliable indicator of inter-rater agreement, namely the coefficient *T*. It was proposed by Tinsley and Weiss (1975) and is described in T. Rietveld and Hout (1993). The coefficient *T* requires that *all* judges agree about each item, and as such it does not take covariance into consideration, and is insensitive to lack of variation in scores. The associated test statistic was developed a few years earlier and described in Lawlis and Lu (1972).

The *T* coefficient is recommended in several places in the literature and is used, for example, in Heldner (2001a) for an investigation very similar to mine, and it definitely has certain merits which warrant its use. *T* values are therefore reported where relevant, but I will also point out some problems with this coefficient and the associated test statistic which makes it unsuitable for certain (larger) data sets, *in casu* a (relatively) large number of judges. The proper application of this test begins with the test statistic, which can reveal if the observed agreement is due to chance or

whether it (the null hypothesis) can be rejected. The test, a genuine χ^2 test, has the following formula (Lawlis and Lu 1972):

$$\chi^2 = \frac{(N_1 - Np - .5)^2}{Np} + \frac{(N_2 - N(1-p) - .5)^2}{N(1-p)}$$

N = number of items (words)
N_1 = number of observed agreements
N_2 = number of observed disagreements
p = the probability that the judges agree on an item by chance

The subtraction of 0.5 is a correction for continuity. The calculation of p, when agreement is defined as identical rating:

$$p = (1/n)^{k-1}$$

k = number of raters
n = the number of point on the scale (here 4)

Inspection of the above formulae will result in the following observations: when the number of judges (k) is high, the probability of achieving agreement by chance (p) is low. With ten raters and four scale points p = 0.00000381. And when p is very low one always obtains a very high χ^2 value, even when the raters agree on only a few out of a large number of items. If ten raters agree on one item out of 907, one gets a χ^2 value of 71.26 and if they agree on two items χ^2 = 647.31. It follows that even the least possible agreement (one in 907) will result in a χ^2 value which is significant beyond the 0.01 level. The problem with this χ^2 test is that the expected value for the 'agreement' cell is too low. The general recommendation is that all expected frequencies should be greater than 5 (Siegel and Castellan 1988, Ferguson 1971), and it is particularly problematic here that the expected frequency of N_1 is much lower than the least possible observed frequency (= 1). To counter this we would need at least 1,000,000 observations. Note that a complete lack of agreement, where N_1 = 0, means that the test cannot be applied.

When the test is applied on the present data 369 observed agreements are found, giving a χ^2 value of 39,246,517.81. Needless to say, it is significant beyond the 0.001 level, but it is also a fairly poor estimate of the level of agreement in this test.

The coefficient T does not have quite the same problems, but it is still sensitive to the number of judges. It too uses p as an estimate of the probability that agreement is due to chance, and when this value is very low, it effectively disappears from the equation, even when it is multiplied by the number of observations. What remains is a coefficient which is to all intents and purposes a ratio of agreements to total number of items. This is not problematic in itself – in fact it may be what one is really after. The problem is that it is sensitive to the number of exact agreements in the material, and this number naturally (though not by mathematical necessity)

decreases as the number of raters increases. It is therefore not possible to compare two or more samples with a different number of raters, and it is not possible to compare a subset of raters with the rest of the group except if the groups are exactly equal in size. Furthermore, one may question the reason for requiring total agreement among all raters. While this makes sense with a small number of judges, it seems inappropriate for experiments with a large number of judges. Consider, as an extreme example, a survey conducted over the Internet with two thousand respondents. If for each item only one person responds differently than the other 1,999 the test score will show no agreement. With only ten raters this is not a huge problem, and the material contains 369 cases (of 907) of total agreement, but the extra strict criterion of exact agreement among all raters does seem unnecessary. There is one further problem with the *T* coefficient. In the paper where Tinsley and Weiss propose this coefficient, they claim that it supplements Lawlis and Lu's χ^2 test, which is only an indicator of whether the observed agreement is greater than chance:

> The investigator, however, also should be concerned with whether inter-rater agreement is high, moderate, or low, not only with whether it is better than chance (Tinsley and Weiss 1975).

Yet, they do not suggest how one should evaluate the *T* score with regard to these three categories, or how one should compensate for the influence of number of raters. It must be stated, though, that any positive value is an indication that the agreement is greater than chance.

As a consequence of the problems outlined above, the *T* coefficient and associated test statistic will be used cautiously, and mostly in contexts where direct comparison between different *T* coefficients is possible, that is, when comparing subgroups of raters of the same size. The primary measure of agreement will therefore be the pairwise comparisons, as reported earlier.

Reporting agreement as a ratio or percentage of pairwise agreements to total number of pairs does not include information about the *magnitude* of the disagreement, but it is reasonable to assume that if two raters give the scores 1 and 3 for an item, then they disagree more than if they had given the scores 1 and 2, or 2 and 3. This information can be captured by the standard measure of *standard deviation*, computed from the raw prominence ratings. The random standard deviation on this scale from 0 to 3 is 1.10, and such high values for individual items may indicate either uncertainty about the item or, especially with very high values above the random score, that the item has been heard in categorically different ways by the raters.

3.4.7 The observed reliability

The results of the reliability test can be seen in Table 3.2.

Table 3.2. Reliability coefficient (Cronbach's alpha) for a group of raters and for a single rater based on ten Danish raters.

		Reliability
Group of raters	$R_{k(f)}$	0.987
Single rater	$R_{1(f)}$	0.886

Both values are high, with the coefficient for the whole group of raters being close to 1. There is hence no doubt that the data as a whole are reliable, that the responses from the ten raters covary. We can then turn to the question of whether the raters agree on the prominence judgements in absolute values.

3.4.8 The observed agreement

A simple overview of the inter-rater agreement can be obtained from the distribution of scores for all ten judges, presented in Table 3.3.

		Ratings (%)			
	Rater	0	1	2	3
Group 1	*r* 1	31	6	43	20
	r 2	31	14	40	15
	r 3	30	17	39	14
Group 2	*r* 4	31	7	47	16
	r 5	31	29	27	14
	r 6	32	0	48	20
Group 3	*r* 7	27	15	44	15
	r 8	31	1	54	14
	r 9	42	18	28	12
Author	*r* 10	31	20	32	17
	Mean	31.7	12.7	40.2	15.7
	S.d.	3.86	9.12	8.89	2.63

Table 3.3. Distribution of ratings as a percentage of the total number of ratings (907). See Section 3.4.5 for an explanation of the values used to represent degrees of prominence and Section 3.4.1 for a description of the rater groups.

There is good agreement about the proportion of words perceived as unstressed (value = 0), ranging from 27% to 32% for all listeners except *r* 9. If this one score is left out the standard deviation is 1.42 instead of 3.86, with a mean value of 30.6%. There is also good agreement about the proportion of 'strong stress' scores (3), with a mean value of 15.7% and a standard deviation of 2.63, and no really clear outliers.

The picture becomes much less clear when we look at the scores for normal stress (2) and weaker stress (1). The notion of weaker stress in particular seems to be problematic for the listeners, with great variation in general and some very obvious

outliers, namely *r* 6, *r* 8, and to some extent *r* 1 and *r* 4. Rater *r* 6 did not perceive a single word as having weaker stress, and *r* 8 commented that he had given up trying to make a distinction between normal stress and weaker stress. It might be questioned whether these raters have then performed the task in a satisfactory manner, but since there was no requirement in the test instructions that all categories be used, this is no ground for dismissing the data. The problem can only be one of interpretation. The category 'normal stress' also shows great variation – probably as a direct result of the uncertainty about the weaker stress category. There is a strong tendency for categories 1 and 2 to be complementary: the more 'weaker stress' responses (1) the fewer 'normal stress' responses (2). The share of *either* 1 or 2 responses varies between 46% and 59% for the ten raters, and the differences between them seem to reflect their use of the scale: *r* 7 has the largest share of '1+2' responses and also the smallest share of 0 responses; *r* 9 has the smallest share of '1+2' responses and the largest share of 0 responses. The first tentative conclusion must then be that whereas there is good agreement about the categories completely unstressed and strong stress, there is much less agreement about the categories normal and weaker stress. Of course Table 3.3 only reveals the number of ratings for each category and not whether the listeners assigned these values to the same words. One way to address this issue is to create distribution matrices (or contingency tables) for each pair of listeners, showing how their scores correlate.

Table 3.4. Distribution matrix for two raters – *r* 1 and *r* 2. Numbers in bold indicate agreement about a specific item/word.

Pairwise comparison				
	r 2			
r 1	*0*	*1*	*2*	*3*
0	**280**	1		
1	3	**34**	13	
2	1	89	**292**	9
3		1	59	**125**

Although distribution matrices like the example in Table 3.4 provide a good overview of agreement and disagreement between two listeners, they are somewhat impractical when the number of listeners becomes too large. With ten raters the number of possible pairs is 45, which makes it difficult to get a full overview of the general tendencies, but the matrices can still be used to check the assumption derived from Table 3.3 – that disagreement is largest between categories 1 and 2 and that these categories tend to be complementary. Table 3.5 shows a matrix of all the responses pooled together, that is, the sums of all 45 pairwise comparisons.[3]

[3] A full list of distribution matrices is available on the accompanying website (see page *vi*).

Distribution of scores – Danish raters				
	y			
x	*0*	*1*	*2*	*3*
0	**30**	1	0	0
1	1	**4**	6	0
2	1	8	**29**	3
3	0	0	4	**12**

Table 3.5. Distribution matrix of all responses in the listening experiment as a percentage of the total number of comparison pairs: 45 rater pairs × 907 words = 40815 pairs.

One of the factors one may look for is whether the disagreements lie between categories 0 and 1 or between 1 and 2. A simple check of the relevant intersections (0–1, 1–0, 1–2, 2–1) for each pair it can be seen that the 1–2/2–1 intersections almost always show a higher value, also reflected in the total scores in Table 3.5, indicating that this is where the disagreement between the raters is strongest.

Pairwise comparison and *T* coefficient

The agreement tests which were described in Section 3.4.6.2 were all performed on the data material as a whole. The results can be seen in Table 3.6.

Pairwise comparisons	
Agreement (mean, %)	75.1
Total agreement	
S.d.	0.29
T	0.41
χ^2	39,246,517
N (no. of observations)	907
N_1 (no. of total agreements)	369
S.d. indicates mean value for the whole material	

Table 3.6. Agreement measurements for the responses of the ten Danish listeners.

The pairwise comparisons for all 907 words – a total of 40815 pairs – yield an average agreement score of 75.1% (grand mean). (By excluding *r* 9, whose responses were aberrant with regard to share of 0 responses in relation to '1+2' responses, this figure rises to 77.2%.) Silverman *et al.* (1992) recommend a criterion of at least 80% agreement for evaluating ToBI transcriptions, but it is very difficult to compare figures across different investigations. This score is of course sensitive to the degree of difficulty of the task and in particular to the number of scale points used. It is therefore

necessary that the tasks be very similar in design if the obtained results are to be compared.

80% agreement is exactly what one gets if only one person out of ten disagrees about the rating, and if two people disagree with the rest the resulting score is 64%. An average agreement score of 80% can therefore be said (roughly) to reflect a situation where on average one person has a different perception of the prominence level of a particular word, which certainly must be said to be a rigid criterion. A result that falls slightly below this, as in the present investigation, still shows very strong agreement overall. Heldner (2001a), in a very similar task (four prominence levels), found 78% agreement within a group of expert raters, and 69% agreement in a group of non-expert raters on the read speech part of his material. These figures are very close to my findings, considering that three of my raters only had fairly limited experience with this type of task. It may also be noted that the agreement level reported in Silverman *et al.* (1992) for *type of pitch accent* was 64% – well below the preferred 80%.

As mentioned earlier it has not been possible to find information about how high T should ideally be, but 0.41 is definitely satisfactory, considering that any positive value shows a correlation above chance. It is comparable to the findings in Heldner (2001a), where T was 0.45 for the prominence rating in read speech. This was based on nine raters, which leads to higher values. The average for any nine raters in my experiment is T = 0.43. Pairwise agreement for the ten possible groups of nine raters varies between 74.2% and 77.2, with an average of 75.1%. Note that the average pairwise agreement for any nine raters is the same as the pairwise agreement between all ten raters, which demonstrates that this method is not sensitive to the number of raters in a group. The group of nine raters which had the highest pairwise agreement is the one which excludes *r* 9, which is confirmation that his responses deviate the most in terms of exact agreement. The χ^2 value is extremely high, which does show that the agreement is highly significant, but it is not a good expression of the degree of agreement, cf. the discussion in Section 3.4.6.2.

The conclusion at this point must be that the data material is highly reliable, and that the ten raters generally agree about the assignment of prominence levels, although *r* 9 may have used the rating scale in a systematically different way (generally lower ratings).

3.4.8.1 Where do the raters disagree?

While the overall reliability and agreement is comfortable, it may also be interesting to look more closely at the disagreements, which will be a further indication of any problems with the perception of prominence or the assumptions of the raters about what constitutes normal stress, weaker stress or strong stress.

I have already mentioned that the problematic distinction seemed to be between normal stress and weaker stress (values 2 and 1). Table 3.7 gives further confirmation about this. It takes one rater – *r* 10 – as a starting point and shows the number of agreements for each response type. The table should be read as follows: *r* 10 answered *x* (e.g. '0') N times. In 'N_1' cases all other raters agreed, giving a pairwise agreement

score of '%' and an exact agreement across all raters of 'T'.

r 10	N	N_1	%	T
0	285	238	94.7	0.84
1	182	0	46.2	0.0*
2	290	57	71.3	0.20
3	150	74	80.5	0.49

N = number of observations (words)
N_1 = number of agreements across all raters
% = pairwise agreements (per cent)
* test conditions not fulfilled

Table 3.7. Agreements and disagreements between rater *r* 10 and the other nine raters in the experiment.

Again we see a very high level of agreement about completely unstressed words, with 94.7% of the pairwise agreements and an agreement across all raters of T = .84 (238 of 285 = 84%). The scores are also high for the category strong stress, but less so for normal stress. And for weaker stress we only see a pairwise agreement of 46.2%, which is only somewhat better than the chance score of 25%. There are no total agreements for this category, which was predictable since one rater – *r* 6 – did not mark any words as having weaker stress. The T coefficient is not an appropriate measure of agreement in such a case, and cannot, or should not, be calculated.

The by now fairly well established pattern becomes even clearer from Table 3.8, which follows up on the distribution matrices in Section 3.4.8. Each possible combination of responses is shown with the number of occurrences and frequency with which it was found. The direction of responses is ignored, so that the pairs where rater *x* answered 0 and rater *y* answered 1 are grouped with the ones where *x* answered 1 and rater *y* answered 0.

The figures in Table 3.8 should be interpreted with an eye on Table 3.3, especially the observed sum for each response type. Agreements about no stress (0), and normal stress (2) account for almost 60% of the pairs as the two most frequently occurring responses and response pairs, and as expected there are also many agreement pairs for strong stress. But when we compare disagreements where one rater scored 2, we find that although there are more 3-ratings (1414) in the material than 1-ratings (1132), there are more than twice as many disagreement pairs between 1 and 2 than between 2 and 3. And the number of disagreements between 0 and 1 is quite low in spite of the fact that the number of 0 scores is fairly high. The number of disagreements involving two scale points (or more) is quite low, but it may be noted that whereas disagreements about 0 and 2 does occur with some frequency, the raters almost never disagree about 1 and 3.

*RP**	*Total*	*Per cent*
0,0	12093	29.6
2,2	11864	29.1
1,2	5896	14.4
3,3	5082	12.5
2,3	2504	6.1
1,1	1629	4.0
0,1	985	2.4
0,2	704	1.7
1,3	49	0.1
0,3	9	0.0
Total	40815	99.9

* Response pattern, rater x=0, y=0, etc.

Table 3.8. Sorted list of the number of occurrences of each possible response pair.

The conclusion about all this must be that while the categories no stress and strong stress are very stable and cause little disagreement, the category weaker stress was treated very differently by the ten raters. The words which were perceived as having weaker stress by some raters were often heard as bearing normal stress, or less frequently no stress, by other raters. There may be several explanations for this. The instructions to the listeners may in some way have made them less disposed towards marking weaker stress, or may have been too vague about what was implied by the term, or it may be because the typical description of weaker stress in Danish is connected with stress reduction in compound words which is not directly paralleled in this experiment. It is also possible that it is not due to any experiment 'error', but that it could instead lead to the formulation of the following hypothesis:

> The difference in the number of 1 (weaker stress) responses and the large number of disagreements about where weaker stress is found is due to the fact that 'weaker stress' is not a linguistically relevant (or at least not a primary) category.

One possible way of testing this hypothesis could be to look at the acoustic manifestation. If it is also difficult to isolate a separate category of reduced or secondary stress based on acoustic properties, that would further support the above hypothesis. But first it would be necessary to see if this tendency is peculiar to Danish listeners, or if it is also found in the ratings of native speakers of British English (see Test 2).

3.4.8.2 Lexical versus grammatical words

In the previous section there was an account of the agreements and disagreements among the raters in relation to the different prominence labels. It was shown that a large number of agreements concern the label 'no stress', that is, almost 30% of the

pairwise comparisons were cases where both raters scored '0' (did not indicate stress above completely unstressed). It turns out on further inspection of the data that all 238 words that all raters agreed were completely unstressed are grammatical words, and *vice versa* that the grammatical words, with a few notable exceptions which will be described later, are perceived as completely unstressed by all raters. One might therefore argue that once this has been established so unequivocally the grammatical words are no longer of any interest. Unless the grammatical words are highlighted as a result of special semantic or pragmatic considerations they are not subject to the same variation in prominence which characterises the lexical words. Section 4.1. In neutral, context-free utterances grammatical words are, as a rule, simply unstressed, and that status has not really been challenged in the present experiment. The situation might be different in spontaneous speech, which has more variation in information structure and other pragmatic/discoursal motivation for (occasionally) placing some degree of prominence on grammatical words. But for the present material one might argue that the (true) agreements and disagreements between raters become clearer if the grammatical words are left out of the analysis.

The overall reliability and agreement scores for the lexical and grammatical words are presented separately in Table 3.9 and Table 3.10. The scores for the whole material are repeated from Section 3.4.8 for comparison.

Reliability				
		Lexical	*Grammatical*	All (repeated)
Group of raters	$R_{k(f)}$	0.946	0.986	0.987
Single rater	$R_{1(f)}$	0.636	0.875	0.886

Table 3.9. Reliability coefficients (Cronbach's alpha) for the lexical and grammatical words in the test separately. Based on ten Danish raters.

Agreement (pairwise and total)			
	Lexical	*Grammatical*	All (repeated)
Pairwise agreement (mean, %)	66.3	93.6	75.1
S.d.	0.39	0.08	0.29
T (total agreement)	0.21	0.81	0.41
N (no. of observations)	612	295	907
N_1 (no. of total agreements)	131	238	369

Table 3.10. Agreement scores for the lexical words in the test. Based on ten Danish raters. The standard deviation value is the grand mean for all words and expresses the average disagreement.

The reliability coefficients are somewhat lower for the lexical words compared with either the whole material or the grammatical words, which are similar, but the most obvious difference is in the agreement scores. Notice especially that the number of total agreements about lexical words is only one third of the number for the whole material although the number of observations is two thirds, that is, the proportion is much lower. This has a serious effect on the *T* coefficient which decreases to half the original value. Conversely, the corresponding numbers for the grammatical words are much higher than for the material as a whole. The numbers demonstrate clearly that the uncontroversial grammatical words have a large, and perhaps undesirable, effect on the overall reliability and agreement scores, so why include them in the analyses, as in the previous section? First, the exclusion of the grammatical words would probably not have a large effect on the inter-relation between raters or groups of raters, only on the size of the difference between them. Secondly, and most importantly, grammatical words are not generally excluded from similar investigations, including some which are used here as a source of comparison, such as Heldner (2001a) and Silverman *et al.* (1992). It can be useful to distinguish between the two groups of words, though, for example in the discussion of the few grammatical words which were perceived as stressed (see Section 6.9).

3.4.8.3 Effect of experience and background

One of the questions which were asked at the beginning of this section was whether the level of experience and the professional background of the raters have a systematic influence on their ratings. In this section I deal exclusively with whether the level of experience affects the degree to which raters agree. The issue of how their background influenced their perception of prominence levels is treated in Section 4.4.

The three groups of raters are described in Section 3.4.1 as two groups of professional phoneticians and a group of students. The hypothesis is that there will be a greater level of inter-rater agreement within the two groups of phoneticians than within the group of students. As a first step the inter-rater reliability is calculated for each group. The differences between the reliability coefficients can be assessed by calculating the test statistic *M*, which is described in T. Rietveld and Hout (1993: 204-205). The procedure was originally developed by Hakstian and Whalen (1976). The test compares the reliability coefficients for a whole group – Cronbach's alpha, or $R_{k(f)}$ – and the significance level of *M* can be seen from a χ^2 table with *df* equal to the number of coefficients – 1. The reliability coefficients for all three groups of raters and the associated test statistic are presented in Table 3.11.

Group	$R_{k(f)}$	$R_{1(f)}$
1	0.969	0.914
2	0.965	0.901
3	0.935	0.829
M	0.23	
df	2	
p	> 0.5	

Table 3.11. Reliability coefficients for the three groups of Danish raters and associated test statistic *M*. The difference between the coefficients is not significant.

The reliability coefficients for all groups are high (close to 1) and very similar, and the test statistic shows that the slight difference between them is not significant. The ratings of each group can therefore be said to be reliable and a valid basis for further analysis, but the reliability coefficients do not indicate the level of *agreement* within each group and any differences between them.

The *T* coefficient is used as a measure of this intra- and inter-group agreement, supplemented by pairwise comparisons. For this specific purpose it is not unreasonable to require total agreement between all raters in the group, and the three groups have an equal number of raters, which allows for inter-group comparisons. However, it is problematic to compare with the results for the entire group of ten raters (see Section 3.4.6.2) so in order to establish a good foundation for the comparison all 120 possible groups of three raters were tested. The results were ranked according to *T* score, and the five highest and five lowest ranking groups are shown in Table 3.12.

Rank	*Raters*	*% (pair)*	*T score*
1	*r*1 *r*8 *r*4	86.6	0.7907
2	*r*8 *r*4 *r*6	85.1	0.7718
3	*r*1 *r*4 *r*6	85.3	0.7671
4	*r*1 *r*8 *r*6	84.7	0.7601
5	*r*2 *r*8 *r*4	83.2	0.7413
...			
116	*r*7 *r*5 *r*6	69.6	0.5190
117	*r*1 *r*7 *r*9	68.2	0.5178
118	*r*1 *r*5 *r*9	67.2	0.5143
119	*r*7 *r*9 *r*6	67.0	0.4978
120	*r*5 *r*9 *r*6	65.0	0.4920

Table 3.12. Five highest and lowest ranking groups of three raters, ranked by *T* score.

None of the three 'natural' groups are represented in the above table, and in fact the

highest ranking group contains one member from each of the three natural groups. The *T* score values of all 120 groups have the following summary statistics:

Minimum: 0.492
Maximum: 0.791
Mean: 0.615
Median: 0.607

Table 3.13 shows the scores for the three 'natural' groups, along with their ranking.

Table 3.13. Scores of the three 'natural' groups and their rank among all 120 possible groups of three raters.

Group	%Agr	T score	Rank
1	78.8	0.664	27
2	73.2	0.578	81
3	70.0	0.545	103

Only group 1 (phoneticians, general and Danish phonetics) agree better than average for a group. Group 2 falls just below average, while group 3 is placed in the lowest quartile.

These observations must lead to somewhat tentative conclusions. The level of experience of the raters does seem to have a possible effect, albeit a slight one, on the level of intergroup agreement, but the effect is not strong enough to warrant any special considerations in the further treatment of the data. The highest ranked natural group barely managed to reach the 75th percentile, while the other group of experienced phoneticians show less intergroup agreement than would be expected for a randomly selected group.

3.4.9 Conclusion

The data material was shown to be reliable with a high degree of overall agreement. The disagreements which do exist mainly seem to be concerned with the use of the category 'weaker stress', but it could not be shown that this source of uncertainty was linked to the raters' linguistic background or level of training. Disagreements about the use of this category was found within all three rater groups. Except for a slight tendency for group 1 (phoneticians, Danish/general linguistics) to agree more than any randomly selected group of three raters, there was no clear effect of level of experience on the consistency within a group.

It was clear from the analysis of agreement and disagreement between the raters that there is considerable variability in the perception of stress levels, especially with regard to syllables or words which are not fully stressed (were not consistently marked as having normal full stress). This means that even minor reduction in the stress or prominence level of a syllable due to syntactic or semantic/pragmatic (or other) factors is likely to be interpreted differently by different raters – even within a fairly homogeneous group. This speaks against the view that a representative rater

can be used as a source of information about larger portions of the material and rather points to the necessity, or at least advantage, of having the entire material assessed by multiple listeners, since their cumulative scores can be used to level out some of the seemingly categorical differences between individual listeners. This notion gains further support from the analysis of perceived prominence levels presented in Chapter 4.

CHAPTER 4

Perceived prominence levels in utterances – Danish listeners

4.1 Selecting and grouping utterances

The underlying structure or idea of the sentence material in this investigation is that some sentences are uttered in a neutral, context-free (default) manner, while others have a marked information structure with one element focused. In order to examine the effect of this on the perception of prominence levels it will be helpful to examine each sentence in a particular context as a whole – in other words to pool the utterances spoken by different speakers of the same sentence in the same context, so that the sentence 'Paul sings' is represented by ten ratings of six different utterances (= six speakers) in identical contexts. That can only be done if there are no clear differences between the realisations of the utterances that are to be grouped. The sample utterance in Table 4.1 illustrates the general pattern:[1]

Spk	*Utterance*
1F:	$^{2.0}$Bill $^{2.1}$struck $^{2.0}$Ann
2F:	$^{2.1}$Bill $^{1.9}$struck $^{2.0}$Ann
3F:	$^{3.0}$Bill $^{1.3}$struck $^{2.2}$Ann
4M:	$^{2.3}$Bill $^{1.7}$struck $^{2.1}$Ann
6M:	$^{2.0}$Bill $^{1.5}$struck $^{2.6}$Ann
*5M:	$^{2.1}$Bill $^{0.5}$struck $^{2.0}$Ann
5M:	$^{1.9}$Bill $^{1.3}$struck $^{2.5}$Ann

Table 4.1. Mean scores from ten raters for the neutral, context-free version of 'Bill struck Ann' (*bsa*). The starred line represents an alternative version from speaker 5M, which was clearly different from, and thus not 'perceptually equivalent' to, the other version.

There are three main points to be observed about this comparison. First, the starred utterance by speaker 5M was an 'alternative' version which was judged by me to be different from, that is not 'perceptually equivalent' to, the other version by the same

[1] Results for all utterances are available on the web page.

speaker, nor indeed to the utterances by the other speakers. This perception is supported by the prominence assignments in this experiment, where 'struck' was clearly deemed to be less prominent in the alternative version than in the 'regular' version. There are three utterances in the material which were realised in two systematically different ways by speaker 5M or 4M, and with enough repetitions of both versions to allow for quantitative analysis. The alternative versions were produced with a different accentual and/or phrasal pattern. They were accepted for further analysis because these alternative versions can provide evidence about how the perceptual differences are achieved acoustically, but they will be excluded from the calculation of means across speakers because they do not represent the 'same' utterance as the others. In all three cases the speaker had also produced another version which was similar to those of the other speakers. Three other utterances had to be excluded, although there were no alternative versions, because their phrasal structure differed from that of the other speakers, or because they were said with emphasis on a particular constituent.

The second point concerns a more systematic difference between the group of speakers as a whole and speaker 3F. The first stressed word of her neutral utterances was often heard as being very prominent – often as prominent as in the utterances where this word was focused. The perceived prominence level of the word, 'Bill', varies between 1.9 and 2.3 for five of the six speakers, but for speaker 3F it was 3.0, indicating that all 10 raters heard this word as having strong stress (score 3). A comparison between speaker 3F's neutral version of *bsa* and the version where 'Bill' is foregrounded shows little difference:

Neutral 3F: $^{3.0}$Bill $^{1.3}$struck $^{2.2}$Ann
Focused 3F: $^{2.9}$Bill $^{1.3}$struck $^{1.8}$Ann

In fact, 'Bill' is perceived as (marginally) less prominent in the version with a marked information structure, but the only clear difference is that the final word, 'Ann', is less prominent in this version. The pattern is the same for all the sentences in Group 1 (see Section 2.2.1), so the ratings for speaker 3F's neutral versions are not included in the mean scores.

Thirdly, one may notice some variation in the prominence level of the last stressed word, 'Ann'. For four speakers the prominence level varies between 2.0 and 2.2, but for speakers 6M and 5M it is 2.5 and 2.6 respectively. These utterances were judged by me to be perceptually equivalent and the latter two are not excluded from the analysis, but it is an important observation which may point to the existence of several valid strategies for realising neutral utterances (further commented on below).

The three observations, or reservations, I have just described do not detract from the overall impression, which is that the utterances are sufficiently similar to be grouped together. The point of grouping the individual utterances is partly that they will then represent a 'prototypical' utterance, but also that each measure will be

made up by enough data points to test the statistical significance of even quite small differences. Below I present only the means across all speakers that can reasonably be grouped together.

4.2 Context-free utterances

The result for the neutral, context-free utterances averaged across all speakers were as follows:

Abbrev.	*Sentence*
ps	$^{2.06}$Paul $^{2.06}$sings
bsa	$^{2.06}$Bill $^{1.70}$struck $^{2.24}$Ann
jkft	$^{2.02}$Jane $^{1.76}$kissed $^{1.98}$Frank $^{2.08}$tenderly
pc	^{0}The $^{2.08}$party $^{0.08}$was $^{2.14}$cancelled
css	^{0}The $^{2.16}$cook $^{0.06}$was $^{1.84}$smelling ^{0}the $^{2.02}$soup
sepc	$^{2.02}$Sheila $^{1.84}$examined ^{0}the $^{1.84}$patient $^{2.10}$carefully
tgios	^{0}The $^{2.08}$Germans' $^{1.74}$import ^{0}of $^{1.96}$sinks ^{0}from $^{2.18}$Denmark
gitsd	^{0}The $^{2.08}$Germans $^{1.72}$import $^{0.02}$their $^{2.02}$sinks ^{0}from $^{2.28}$Denmark
pdp	$^{0.57}$Is $^{2.00}$Peter ^{0}a $^{1.77}$doctor ^{0}in $^{2.23}$Paris
dsi	$^{1.50}$Did $^{2.00}$Stalin $^{1.70}$insist ^{0}on ^{0}an $^{1.82}$equal $^{1.67}$distribution ^{0}of $^{2.17}$wealth

The figures represent mean scores across five speakers (3F excluded, see above) by all ten raters, so that each score is based on 50 observations. Sentences *pdp* and *dsi* are only represented by three and four speakers respectively, giving 30 and 40 observations for each of these sentences. As can be seen from the prominence values, grammatical words were in general, and as expected, perceived as completely unstressed (see also Section 3.4.8.2). Apart from one rater's judgement of the word 'was' as carrying weaker stress there are two more significant exceptions from this pattern, namely the words 'Is' and 'Did', which are initial in sentences *pdp* and *dsi* respectively. There is a very large degree of disagreement among the raters about the prominence level of these items, and for this and other reasons I will argue that these words are best treated as a type of 'high prehead', in which the status of their accentuation (are they stressed/accented or not?) is perhaps indeterminable. In the following analyses (Tests 1–3) the first fully stressed, or accented, word is assumed to be the first lexical item of the sentence ('Peter' and 'Stalin' in *pdp* and *dsi* respectively). The issue of these high preheads is treated in some detail in Section 6.9. Since the grammatical words are always unstressed they can be excluded from further analysis. The ratings for the lexical words are depicted graphically in Figure 4.1.

Two major trends appear from the diagrams in Figure 4.1. (1), the first and last lexical items in a sentence are generally deemed to be the most prominent, and (2), the prominence levels on these and the intervening stressed words seem to follow a strong – weak alternating pattern.

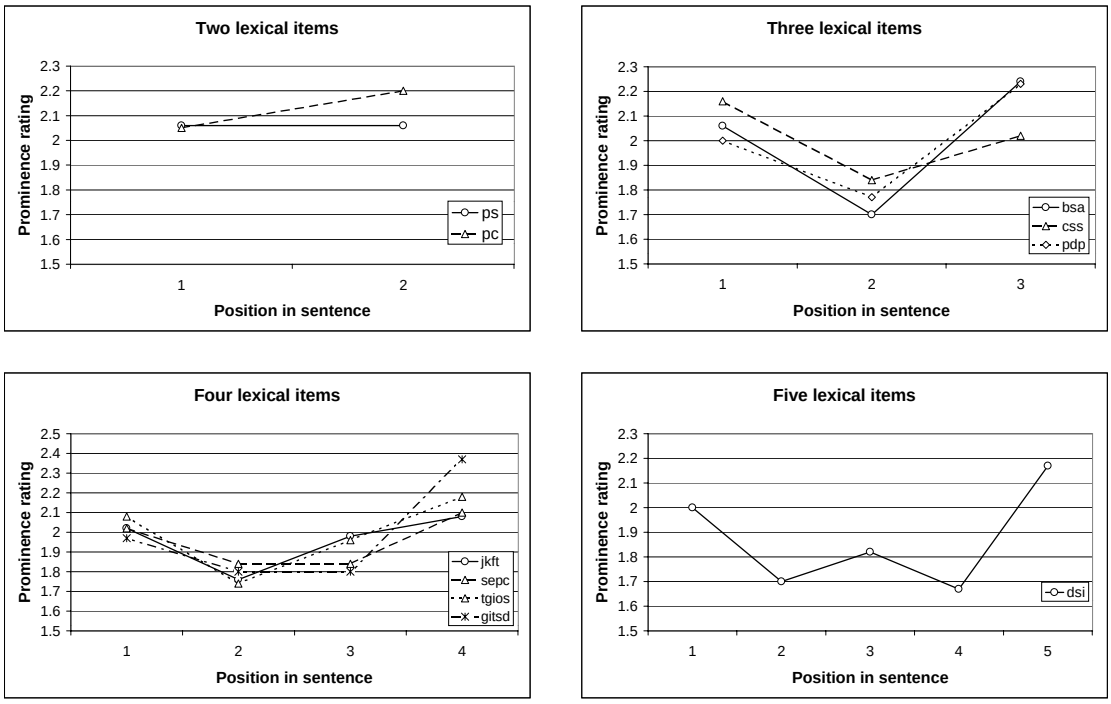


Figure 4.1. Stressed syllables in neutral sentences. Means across all speakers (excluding 3F) and all raters.

4.2.1 First and last lexical item – or onset and nucleus

Most traditional descriptions of intonation in English rely on the concept of an *intonation unit*, or *tone group*, which has a regular structure. The first stressed syllable of an intonation unit is sometimes referred to as the *onset*, (e.g. in Knowles 1987) and the last stressed syllable is usually called the *nucleus*. An utterance can consist of several intonation units, and ideally the boundaries between them are marked by pauses, reduced speech rate or by other phonetic means. This is referred to as 'external' evidence in Cruttenden (1997). In the absence of clear, external evidence one may sometimes have to rely on 'internal' evidence, which means constraints set by the obligatory structure which has been established for an intonation unit. For example, if one recognises two nuclei in a stretch of speech then there must necessarily be two intonation units, since an intonation unit must contain one and only one nucleus. This circularity involved in assigning intonation unit boundaries is sometimes noted in the literature (Cruttenden 1997: 29) but it is not acceptable in this study, since the very notion of the nucleus and its relation to prominence is under investigation. I have only accepted the presence of a boundary if there was clear rhythmical evidence for it, perhaps in connection with pitch cues. Such clear boundaries were infrequent in my material of short utterances; even the sentence 'Did Stalin insist on an equal distribution of wealth', with five stressed words, was typically realised without any clear boundaries. Only the sentences with a final adverbial – 'carefully, tenderly' –

were regularly split into two intonation units, or phrases, and there was almost always another, more frequent, version without a boundary by the same speaker. Since the few utterances with phrasing into several intonation units were excluded from analysis (see the 'alternative versions' above), it follows that, according to my definitions, all the utterances consist of one intonation unit. As such, the stressed syllable in the first lexical item, and therefore fully stressed word, in each utterance should be equivalent to an onset, and the last lexical item is expected to be the nucleus. It cannot, however, be assumed that an analysis or transcription of the utterances according to the British school of intonation analysis would yield the exact same result, mostly because the criteria for locating boundaries within the British tradition are not identical to mine. A further indication of the lack of clear boundaries is that the raters did not consistently perceive any boundaries in any of these utterances. I therefore assume in the following presentation that the first and last stressed syllables in an utterance, found in the first and last lexical items, are equal to onset and nucleus respectively.

The prominence ratings for both the first and last lexical item place them at or slightly above the category of normal stress. The ratings vary between 2.00 and 2.28 and in no sentence is there an average difference of more than 0.23 points on the prominence scale. The small differences in ratings which form the basis for the analysis below of course only indirectly represent differences in rater perception. If one word received the score 2.0 and another the score 2.2 it does not, strictly speaking, mean that the (individual) raters deemed the second word slightly more prominent. They could only choose between four labels, which have been translated into discrete numerical values. A difference in mean score therefore reflects the number of raters which deemed one word to be more prominent than another word. But it is my contention that, given the high number of observations (around 30-60) for each word, the small differences in ratings can be taken as evidence of small differences in perceived prominence, and a difference of 0.1 is consequently assumed to indicate that one word is slightly more prominent than another. The observed differences are naturally tested statistically.

The differences between first and last lexical item are small and go in both directions, but there does seem to be a tendency for the last item to be deemed slightly more prominent. This was tested statistically (t-test) for each of the sentences and for all ten sentences grouped. The results are shown in Table 4.2.

There is an overall difference of 0.10 degree of prominence between the first and last items, and the difference is highly significant ($p < 0.001$) due to the large number of observations: 440 pairs. This seems to support the idea that the last stressed word, the nucleus, is the most prominent one in the intonation unit, but such a general and strong conclusion would be too simplistic. The difference is only significant in four out of the ten tested sentences, and in one sentence the first item was deemed more prominent than the last (the difference just fails to be significant at the 0.05 level). When we look at the individual utterances it is clear that there is quite a large

Sent.	*First*	*Last*	*Diff.*	*N*	*p*
ps	2.06	2.06	0.00	50	1.000
bsa	2.06	2.24	0.18	50	**0.028**
jkft	2.02	2.08	0.06	50	0.261
pc	2.05	2.20	0.15	40	0.057
css	2.16	2.02	-0.14	50	0.070
sepc	2.02	2.10	0.08	50	0.159
tgios	2.08	2.18	0.10	50	0.200
gitsd	1.97	2.37	0.40	30	**0.000**
pdp	2.00	2.23	0.23	30	**0.006**
dsi	2.00	2.17	0.17	40	**0.033**
All	2.05	2.15	0.10	440	**0.000**

p = two-tailed probability, paired t-test
Significant values ($p < 0.05$) are in bold-face type.
N = number of pairs

Table 4.2. Prominence levels of first and last lexical item in the context-free sentences. The difference 'last – first item' is listed in column four.

spread in the differences between the two items. Figure 4.2 shows the distribution of ratings of the first and last item and the differences between them for all 44 neutral utterances.

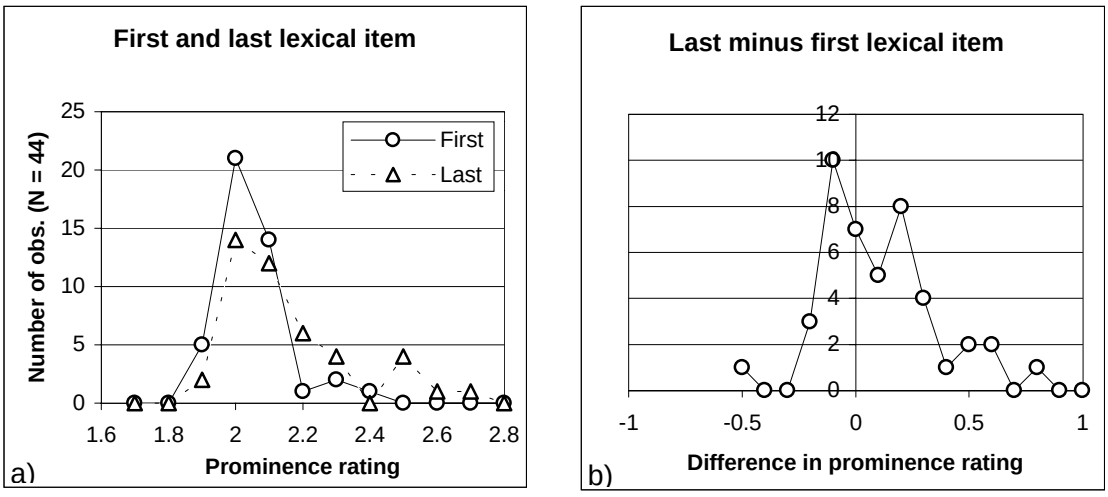


Figure 4.2. Prominence ratings for first and last lexical item (a), and the differences last minus first item (b).

In most of the utterances the difference is 0.3 degree of prominence or less, and none of these differences is statistically significant. In six utterances the last lexical item is significantly more prominent than the first ($p < 0.05$), and in one utterance there is a

strong tendency for the first item to be more prominent than the last ($p = 0.052$). It thus seems that the general tendency for the last lexical item to be slightly more prominent is partly, or mainly, caused by a small number of utterances where this is very clear. The six individual utterances, that is, not grouped across speakers, where the last item is more prominent than the first are:

Sentence	*Speaker*	*Difference*
gitsd	2F	0.8
bsa	6M	0.6
bsa	5M	0.6
dsi	1F	0.5
tgios	6M	0.5
sepc	6M	0.4

There is no obvious pattern in terms of which sentences are affected. With regard to speaker effect it can be noted that speaker 6M uttered three of the six utterances, but he also produced the only utterance with a noticeably more prominent first lexical item (*css*). It is very clear, though, that there is a strong correlation between these six utterances and the sentences with an overall significant difference (see Table 4.2); the four utterances with the largest difference come from three of the four groups with an overall significant difference. The fact that the difference in prominence ratings between first and last lexical item stems largely from a few utterances seems to imply that this is a question of choice of strategy rather than a general rule: sometimes the last syllable of a neutral utterance is clearly made more prominent than the other words in the utterance, but most of the time it is not.

Two of the clearest overall differences are found in the sentences *pdp* and *dsi*, which are both yes/no-questions. One of the speakers has a falling tone on the last lexical item of these utterances, while the other two and three speakers, respectively, have a rising tone. There is a slight tendency for items with a rising tone to be deemed more prominent than those with a falling tone, and it may be that the rising pitch is more 'prominence lending' than the falling pitch found in the other sentences. However, there are too few occurrences in my material to test this hypothesis, and the variation may simply be random.

The ten raters are fairly uniform in their ratings of first and last lexical item, but one rater had a slightly stronger inclination towards a higher score on the last item, as can be seen in Table 4.3.

For nine of the ten raters the difference is smaller than 0.2 degree of prominence, but for rater *r* 1 it is 0.39. This rater's scores do substantially influence the mean scores of the whole group; if her scores are excluded the overall tendency becomes much weaker, and if the *three* raters with the largest difference are excluded the difference between first and last lexical item is only statistically significant for two sentences and not significant overall.

Spk	*r* 1	*r* 2	*r* 3	*r* 4	*r* 5	*r* 6	*r* 7	*r* 8	*r* 9	*r* 10
First	2.05	1.98	2.02	2.02	1.98	2.39	2.05	2.05	1.89	2.07
Last	2.43	2.14	2.11	2.02	2.05	2.48	2.14	2.00	1.98	2.18
Diff	0.39	0.16	0.09	0.00	0.07	0.09	0.09	-0.05	0.09	0.11
p	**0.00**	**0.01**	0.10	1.000	0.37	0.52	0.16	0.16	0.21	0.06
p = two-tailed probability, paired t-test										

Table 4.3. Mean scores of first and last lexical item for each of the ten raters.

The conclusion must be that there *is* indeed a tendency for the last stressed word in the neutral utterances to be deemed slightly more prominent than the first stressed word; the difference is found in *most* of the individual utterances and for *most* of the speakers, and therefore seems to be fairly stable. Yet, it is important to realise that the differences are very small in almost all utterances and for almost all speakers, and the only reason that the difference was statistically significant in even four out of ten sentences was that there was a clear and strong difference in a few utterances representing these sentences.

It is therefore not justifiable, in my opinion, to define the nucleus, which is normally considered an obligatory element of the intonation unit, as the most *prominent* syllable (or word) in this unit. It may be, and as other utterances in my experiment will show in following sections often is, but in neutral, context-free utterances this description is unsatisfactory.

It is of course possible to question the whole idea of an obligatory, default nucleus in all utterances, or intonation units, but even if we accept this idea of a compulsory nucleus it is important to realise that it need not be connected with the notion of prominence. The data in this study support the idea that it is perfectly possible to have a nucleus which is not the most prominent syllable in the intonation unit but which can still be identified as the nucleus based on its role in marking the information structure of the utterance. The issue of nucleus identification has been a matter of some controversy, and especially Brown *et al.* (1980) have challenged the traditional view that nuclei are fairly easy to identify. Their work builds on experiments carried out by Karen Currie and reported in Currie (1978) and Currie (1979). In the first experiment reported in Brown *et al.* (1980) listeners were asked to identify the tonic (nucleus) in a series of read sentences. They were allowed to mark as many tonics as they wished in each sentence. The sentences had been analysed instrumentally and a set of three phonetic cues had been established for each sentence: A (maximum pitch height), B (maximum pitch movement) and C (maximum intensity). They found great variation in tonic identification among the 29 judges. Even the cases where all three phonetic maxima were found on the same syllable gave rise to some disagreement, and the single strongest factor in tonic perception was found not to be any (or all) of these maxima at all. In their list of conclusions the first item

reads:

> Judges will choose the last lexical item as tonic if there is no strong phonetic competition elsewhere in the sentence (nine sentences). They will sometimes choose the last lexical item if all the phonetic maxima are located elsewhere in the sentence. It appears, then, that the last lexical item is regarded as being the tonic *by right of being the last lexical item* if some other item is not heavily marked phonetically as being in competition. (Brown *et al.* 1980: 145-146)

So, in their study there was often no simple correlation between phonetic cues and the identification of the tonic. And the phonetic cues they identified are those which have traditionally been associated with stress, or prominence (Fry 1958a). This further supports the idea that the perception of prominence and identification of the nucleus should be viewed as separate, but obviously interrelated phenomena.

The conclusion proposed above about the connection between prominence and nucleus identification naturally cannot be generalised beyond the rather restricted type of material which was analysed here. In particular, there could be a connection between the perceived prominence of the nuclear syllable and the intonational structure of the utterance. Cruttenden (1981) examined nucleus perception in a number of sentences with systematically varying intonation (following the transcription practice from O'Connor and Arnold 1961), and found that

> there were just [...] two types of IP [= intonational phrase ~ intonation unit] where the word beginning the nuclear tone was not regularly selected as the most prominent. These were high head plus low fall and fall plus rise (personal communication).

The clearest examples were the ones with a high head plus low fall, and although my utterances were not systematically analysed in the tradition of the O'Connor and Arnold system, it is of course possible that many of them would fall in this category. However, this would not in itself provide evidence for dismissing my earlier conclusion. First, most of the other intonation unit types in Cruttenden's experiment involved some kind of advanced nucleus placement, that is, explicitly marking some other word than the last lexical item as the informationally 'most important' word. These sentences are best compared with the sentences with a marked information structure in my investigation and not with the context-free utterances. The only relevant comparison is with Cruttenden's sentences with a *high* fall (on the nucleus) instead of a low, and in those sentences the fall was widely recognised as the nucleus. While this might be partial explanation of my results, it is still not counter-evidence to my conclusion about the relative infrequency of an especially prominent nucleus; even if we could accept a claim that the present material contains a large number of utterances with a low fall nucleus we would still have to explain why this is the case. But such a claim does not seem entirely appropriate. The acoustic data presented in Section 2.5 do not seem to support it: it was generally found that the nucleus was not downstepped, but that F_0 on this item was level with or slightly higher than the

previous stressed syllable. Secondly, the task in Cruttenden's experiment was different than in Test 1. The judges were asked to:

> Please write down the word which you hear as most prominent or standing out as most important in the sentence (Cruttenden 1981: 340).

The judges were asked to select *one* word whereas in Test 1 they were asked to indicate the *degree* of prominence on each word, and were allowed to assign the same level of prominence to several words. Also, the phrase 'standing out as most important in the sentence' is subtly, but perhaps crucially, different from simply 'prominent'. What is considered most important in a sentence may be linked closely to the semantic content, and thereby the information structure, of that sentence, since it must be assumed that most people pay more attention to content than to form. And while the notion of 'prominence' may also be associated with meaning or information structure to some degree, it is part of my point here to show that these two concepts should not be equated. Even if a panel of listeners could unequivocally identify the last stressed syllable in all utterances in the present study as the nucleus it would not mean that the nucleus was more prominent. It should be acknowledged here, though, that Cruttenden does mention the possibility of a nucleus which is not the most prominent syllable in the intonation unit – a view which he also expresses in Cruttenden (1997: 43-44), and which distinguishes him from most, if not all, other proponents of the British school of intonation analysis.

4.2.2 Intervening lexical items: strong – weak alternation

The section above focused on the two peripheral stressed words in the utterance – the first and last lexical items – because these positions seem to be deemed the most prominent by the raters. An observations which agrees with the findings of Widera *et al.* (1997) (see Section 1.4.3). Conversely, that means that any intervening stresses are perceived as less prominent. This tendency is quite clear in all the neutral utterances, as can be seen in Figure 4.1 above. In addition, one other trend appears from the diagrams: the prominence on consecutive lexical items seems to follow a strong – weak alternating pattern.

In sentences with three lexical items there is a clear reduction in perceived prominence level on the middle one, which is significantly less prominent than both the first and last items in all three sentences ($p < 0.05$, two-tailed t-test). In sentences with five lexical items the first and the last item are again clearly deemed the most prominent. The differences between those and any one of the intermediate items are significant ($p < 0.05$), while the differences between the third lexical item and the surrounding slightly more prominent items are not significant, but at least show a tendency towards strong – weak alternation. In sentences with four lexical items the picture is less clear. Again the first and last items are deemed most prominent, and there is a clear drop in prominence level from number one to number two (significant for all sentences except *gitsd*). But the prominence level of the third item varies; in sentence *sepc* and *gitsd* it is as low as number two (and significantly different from the

following, final item) and in *jkft* and *tgios* it is almost as high as the last item (in *tgios* the difference is still significant).

These observations lead to the following hypothesis: the first and last lexical items in a (monophrasal) utterance are always the most prominent, and the prominence levels generally follow a strong – weak alternating pattern from left to right. With an equal number of lexical items in the phrase the penultimate item is in a position of conflict: it should be strong in relation to the preceding item but weak in relation to the following, final item. In this case there seems to be a choice for the speaker to make it either weak(er) or strong(er).

My material is not comprehensive enough to test this hypothesis, but it warrants further investigation. It has been observed elsewhere that in the case of three consecutive stresses in an intonation unit the speaker has the option to deaccent the middle one; or sometimes expressed a little differently: 'any accented syllables between onset and nucleus are liable to lose their accent in all but [s]low deliberate speech' (Knowles 1987: 124). Knowles states that 'accent suppression is not all-or-none; it is a process that can apply to a greater or lesser degree' (p. 126), and indeed, the intervening stresses in these sentences are not fully deaccented. They all have an average prominence rating between 1.6 and 2.1, that is, much closer to the label 'normal stress' than the label 'weaker stress', or indeed 'unstressed'. So the strong – weak alternation in this material is not a question of complete deaccentuation but operates, in my opinion, *within* the category of normal utterance level stress.

There were however a few utterances where the deaccentuation was very clear, cf. the alternative version of *bsa* by speaker 5M mentioned at the beginning of this section, but they were excluded from analysis because they sounded markedly different from the other versions. This is an indication that complete deaccentuation is also an option for the speaker.

4.3 Utterances with marked information structure

The role of the utterances where a marked information structure had been elicited was to investigate how focusing a word affects both the focal stress/accent and the non-focal stresses. The expectation was that the focused word would be more prominent – carry strong stress – and that the non-focal words would be less prominent – carry weaker stress – in relation to the neutral utterances. The initial auditory analysis seemed to confirm this, but as was the case with the neutral utterances there are also some 'alternative' versions of these utterances which tend to break the regular pattern. This concerns the sentences *jkft* and *sepc*, which were sometimes broken up into several phrases by some speakers, which is obviously reflected in the prominence relations: a few raters assigned the label strong stress to non-focal stressed words. These cases are similar to the problem mentioned at the beginning of Section 4.1 and illustrated in Table 4.1, but they are not quite as easy to identify with certainty, especially since there are no 'normal' versions to compare with for the same speaker. Since boundary markings were only used very sparsely by the raters in this experiment, I decided to use the ratings from Test 3 (see below) as a criterion: if at least

three of the four raters in that test agreed that a boundary was present, the utterance was excluded, which affected six utterances: four of *jkft* and two of *sepc*. In all six cases there was a boundary before the final adverbial, and in four cases the word immediately before the boundary was the focused word. The prominence ratings for all the utterances with marked information structure which can be grouped together into structural types are listed below. All numbers represent raters' averages across at least four speakers. The 'focus' labels *f1, f2, f3* and *f4* indicate that the intended focus is on the first, second, third and fourth lexical item respectively.

Focus	*Sentence*
f1	$^{2.95}$Paul $^{1.65}$sings
f2	$^{1.68}$Paul $^{2.90}$sings
f1	$^{2.93}$Bill $^{1.27}$struck $^{1.78}$Ann
f2	$^{1.67}$Bill $^{2.98}$struck $^{1.72}$Ann
f3	$^{1.96}$Bill $^{1.38}$struck $^{2.96}$Ann
f1	$^{2.94}$Jane $^{1.28}$kissed $^{1.54}$Frank $^{1.72}$tenderly
f2	$^{1.67}$Jane $^{2.98}$kissed $^{1.50}$Frank $^{1.60}$tenderly
f3	$^{1.77}$Jane $^{1.40}$kissed $^{2.90}$Frank $^{1.60}$tenderly
f4	$^{1.96}$Jane $^{1.58}$kissed $^{1.78}$Frank $^{2.70}$tenderly
f1	^{0}The $^{2.93}$party $^{0.07}$was $^{1.65}$cancelled
f2	^{0}The $^{1.82}$party $^{0.05}$was $^{2.83}$cancelled
f1	^{0}The $^{2.98}$cook $^{0.07}$was $^{1.53}$smelling ^{0}the $^{1.63}$soup
f2	^{0}The $^{1.70}$cook $^{0.05}$was $^{3.00}$smelling ^{0}the $^{1.65}$soup
f3	^{0}The $^{1.90}$cook $^{0.05}$was $^{1.62}$smelling ^{0}the $^{2.78}$soup
f1	$^{2.98}$Sheila $^{1.44}$examined ^{0}the $^{1.52}$patient $^{1.56}$carefully
f2	$^{1.77}$Sheila $^{3.00}$examined ^{0}the $^{1.48}$patient $^{1.55}$carefully
f3	$^{1.75}$Sheila $^{1.50}$examined ^{0}the $^{2.95}$patient $^{1.60}$carefully
f4	$^{1.90}$Sheila $^{1.65}$examined ^{0}the $^{1.67}$patient $^{2.87}$carefully
f1	$^{0.02}$Is $^{2.68}$Peter ^{0}a $^{1.56}$doctor ^{0}in $^{1.86}$Paris
f1	$^{0.76}$Did $^{2.72}$Stalin $^{1.54}$insist ^{0}on ^{0}an $^{1.62}$equal $^{1.52}$distribution ^{0}of $^{1.94}$wealth

As with the neutral utterances it is clear that the grammatical words, which were expected to be completely unstressed, were indeed judged to be so. The only exception (disregarding one rater's judgement of the word 'was' as having weaker stress) is 'Did' in *dsi*, which represents a special case which is treated in Section 6.9 and disregarded in the following. The unstressed words can therefore be excluded from further analysis. The lexical words, which all achieved an average prominence rating close to 'normal stress' in the neutral utterances, all have an average rating of more than 1.2 in these utterances, so the pre- and post-focal lexical items are not perceived

as completely unstressed, at least not by all raters. Although the general pattern of prominence relations in the utterances can be seen from the raw data, it becomes clearer from a graphic representation as in Figure 4.3.

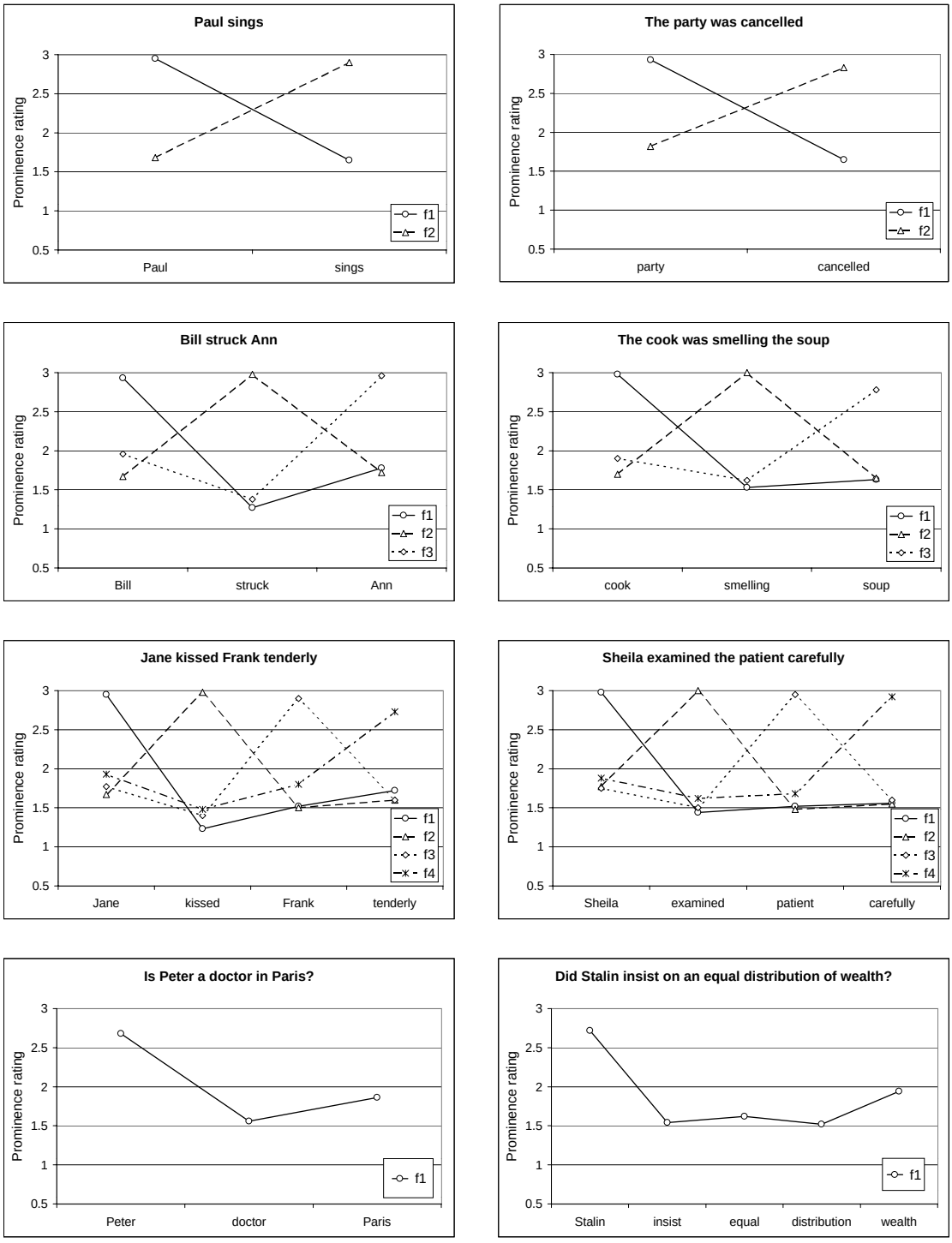


Figure 4.3. Prominence ratings of the stressed syllables in sentences with a marked information structure. The labels *f1, f2, f3, f4* indicate that the position of the intended focus is on the first, second, third and fourth stressed syllable respectively.

4.3.1 Focal stress/accent

It is clear that the word which was expected to be focused was indeed deemed to be the most prominent word by the raters, achieving an average rating of 2.7 or higher in all cases and of 2.9 or higher in 70% of the cases. In other words, the focused words were deemed by almost all raters to carry 'strong stress', which is as expected. The limited variation that does exist does not seem to be strongly systematic, although there is a slight tendency towards a lower prominence rating when the focused word is the last stressed word in the utterance, or penultimate in utterances with four stressed words. The average prominence rating for all focused items in utterance final position is 2.85, and the rating for non-final focused items is 2.93. The difference of 0.08 on the prominence scale is statistically highly significant ($p < 0.001$, two-tailed t-test) but very small.

4.3.2 Non-focal stress

The stressed but non-focal words were deemed much less prominent than the focused words, typically by 1–1.5 degrees of prominence. This is true of both pre-focal and post-focal lexical items. Less obvious and more interesting is the fact that the non-focal items were also deemed less prominent than the same words in the neutral utterances, which shows that these words do not just have reduced prominence relative to the focused words, but also relative to a neutral production with no marked information structure. In the six sentences where information structure is varied systematically, there are 40 words in stressed but non-focal position, and all were deemed less prominent than the corresponding words in the neutral utterances. The differences range from 0.1 to 0.55 degrees of prominence and are statistically significant in 36 cases ($p < .05$); in the remaining four cases the difference just fails to be significant at the 5% level (two-tailed t-test). Although this tendency is very clear and statistically significant, it is, again, worth noting that the difference is on average about one third of one degree of prominence, reflecting the fact that only a minority of the raters assigned the label weaker stress or no stress to these words.

There is systematic variation in the prominence levels of the non-focal lexical items within an utterance, so that items farther away from the focused word are deemed more prominent than those closer to the focal stress. This is reflected (graphically in Figure 4.3) in falling curves from one pre-tonic lexical item to the next in for example *bsa*, *css*, *jkft*, and *sepc* (f3) and in rising curves from one post-tonic word to the next in *bsa*, *css* and *jkft* (f1) and *pdp*. There may be several ways to account for these observations, and one might consider the following two hypotheses: (1) the perceived prominence level of a non-focal stressed word is proportional to its distance from the focused word. Or the less general alternative, (2) the first and the last stressed word of an utterance (or intonation unit) are less reduced in non-focal position than other non-focal stresses. The first hypothesis is the stronger and more general of the two. It gains some support from the above-mentioned sentences, but there are also a number of exceptions to this general pattern. In *jkft* (f4) there is a rise in perceived prominence level from the second to the third pre-focal lexical item, and in

sepc (f4) and *dsi* the non-peripheral, non-focal items are equally prominent. These deviations would have to be accounted for by some other principle or mechanism (or discarded as random errors) in order to maintain the hypothesis. Best support for the hypothesis comes from the fact that in sentences with more than two lexical items – *bsa, css, jkft, sepc* – the prominence level of the first stress is proportional to its distance from the focal word. The differences in prominence level are significant in five out of the eight possible combinations, for example 'Bill' in *bsa* (f2) versus 'Bill' in *bsa* (f3), ($p < 0.05$), and overall when comparing sentences where the focal stress is immediately following with sentences where the distance to the focal stress is maximal ($p < 0.001$). There are traces of this principle in the second non-focal lexical item, too, but much less clearly so. The post-focal items do not seem to follow the principle at all, but are often deemed equally prominent by the raters. There is still some variation in prominence level depending on the position in the utterance, that is, prominence levels increase slightly towards the end of the utterance, but the distance from the focal item does not seem to matter. The post-focal items at for example position three or four in *jkft* and *sepc* are equally prominent regardless of the position of the focused word. It seems clear that the stronger and more general hypothesis can only be sustained if it is combined with some other principles which can account for the obvious deviation from the general pattern.

The second hypothesis was that the two peripheral non-focal items – the first and last lexical item of the intonation unit – are in some way special and less subject to reduction than the other non-focal stresses. The pre-focal initial lexical item in *bsa, css, jkft* and *sepc* all follow this pattern, as do the post-focal, final lexical item in the same sentences and in *pdp* and *dsi* (although the tendency is quite weak in *sepc*). The differences between the first and the immediately following item are significant in five out of six cases ($p < 0.01$). In the last case the tendency is still fairly strong, with $p = 0.054$. The differences between the last and the immediately preceding item are only significant in *bsa, pdp* and *dsi*, and it thus seems that the first stressed word of an utterances varies more as a function of its distance to the focal stress than does the last. Although there are no obvious exceptions to the general pattern, the problem with this hypothesis is that it only accounts for part of the variation that can be observed and is therefore not a very strong one. It too may need to be supplemented by other principles to give a fuller picture of the perceived reduction of the non-focal stresses.

The present material is not comprehensive enough to provide a full and satisfactory explanation of the prominence relations in the non-focal stresses, but the observations suggest the following tentative explanation which can serve as a starting point for further investigation, and which also sums up this section.

(1) In utterances with a marked information structure the focused word is clearly made more prominent than just 'normal stress'. This is true in all positions in the utterance, although there is a (weak) tendency for the focused word to have less perceived prominence in final position.

(2) Non-focal items in such utterances are reduced – both in relation to the focused word in that utterance and relative to 'normal stress'. The non-focal items are, as a first approximation, reduced *as a whole*, that is, they are all reduced a certain amount but the relation between them will still be similar to the relation between the items of the same utterance in a neutral context.

(3) Pre- and post-focal contexts are treated differently, so that pre-focal items are reduced in inverse proportion to their distance from the focal accent, while post-focal items seem to be further compressed. There is some variation in prominence level with position in the utterance, so that a post-focal item early in an utterance is less prominent than an item later in the utterance, but the prominence level does not depend on the position of the focused word. In general, the post-focal items were deemed less prominent than pre-focal items. In sentences that are represented with both pre-focal and post-focal contexts, that is, *ps, pc, bsa, css, jkft, sepc* we find the following values: pre-focal = 1.70, post-focal = 1.58 on the scale from 0 to 3; the difference is significant ($p < 0.001$). A simpler formulation of these observations is that post-focal items are more reduced than pre-focal items and less dependent on the position of the focal accent. The significance of this is tested and discussed in more detail in Chapter 7.

Some of the principles are illustrated in Figure 4.4.

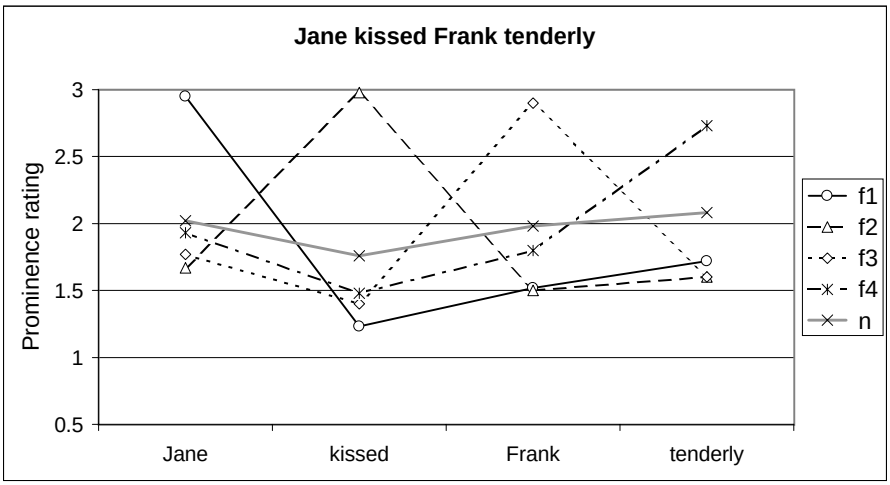


Figure 4.4. Prominence ratings for sentence *jkft*. Both the neutral (n) version and the versions with a marked information structure (*f1-4*) are represented.

The lines connecting the values of the pre-focal lexical items in the various utterances with marked information structure in Figure 4.4 run parallel to each other and to the lines connecting the values of the corresponding lexical items of the neutral version. But the closer the items are to the focused word, the lower the line. The lines connecting the values of the post-focal stresses seem to converge, but still run parallel to the lines connecting the values of the lexical items in the neutral version.

The displacement in prominence ratings in Figure 4.4 might be seen as an artefact of the experimental method, or of general perceptual properties; that when we hear an utterance with a focused word the perceived prominence of the non-focused words are automatically influenced even if there are no local phonetic cues to lead to this reduction. But if this were true we should expect the perceived prominence on pre- and post-focal items to be exactly parallel – and it is not. It must therefore be assumed that the perceived pre- versus post-focal differences reflect real and systematic differences in prominence relations.

4.4 Effect of experience and background on perceived prominence

In Section 3.4.8.3 the connection between agreement between raters and the amount and type of their linguistic experience was examined. Although it was shown that the more experienced raters generally agreed slightly better than inexperienced raters, there did not appear to be much effect of the type of linguistic training or experience. The intra-group variation seemed as large as the inter-group variation, with regard to the general reliability and agreement measurements. This did not rule out the possibility of systematically different perceptions of prominence levels which might be caused by the different linguistic traditions of the raters. In particular, it might be expected that the 'general phoneticians', who mostly work with Danish (Group 1), and perhaps the Danish phonetics students in Group 3, would respond differently from the phoneticians trained in and working with English phonetics (Group 2). The prominence ratings of all three groups are presented graphically in Appendix B, Section B.5. The graphs do not reveal any systematic and consistent differences between the groups. The ratings for the utterances with a marked information structure, or (narrow) focus, are almost identical, while the diagrams for the neutral, context-free utterances show a few isolated differences: Group 3 did not perceive stress reduction on the second item 'smelling' of *css n*; Group 2 heard the final item (the nucleus) of *ps n* as considerably less prominent than the first item, while the other rater groups heard the final item as more prominent; and Group 1 perceived much less reduction on the item between onset and nucleus in sentence *dsi n*. But these are simply random fluctuations which cannot be assumed to be linked to linguistic training. Two observations regarding Group 2 deserve a comment. (1), this group seems to be more sensitive to the reduction of the second lexical item, especially compared with Group 1, but also, for sentences with three lexical items, compared with Group 3. This means that the group of English language phoneticians in this respect resemble the native English listeners more than the other groups (see Chapter 5). (2), there is no tendency for this group to perceive the final lexical item as more prominent than the other groups. It might have been expected that their knowledge of English intonation – in theory and practice – would make them more sensitive to the putative stronger prominence of this item, but this was not the case.

In conclusion, no systematic differences could be detected between the rater groups, but the relatively large variation on selected items points to the need for a

relatively large number of raters in investigations of perceived prominence, in order to level out the idiosyncrasies of individual raters.

4.5 Preliminary conclusions

The prominence levels of the lexical items in the utterances, as perceived by the Danish raters, exhibit some systematic variation, of which only some corresponds to the predictions made by, for example, the traditional British school of intonation analysis. The first and last lexical items in neutral, context-free utterances were generally perceived as the most prominent ones, which to some degree is mirrored in the special status these two positions normally have within the British framework as onset (secondary accent) and nucleus (primary accent) respectively. The general claim that the nucleus is always the most prominent syllable in the utterance, or strictly speaking the intonation unit, could not be substantiated. Although this was sometimes the case, the tendency was really only clear in a minority of the neutral utterances. In the majority of these the difference in perceived prominence between the first and last lexical items was negligible or even in the opposite direction of what might be expected. The perceived prominence of lexical items between the first and last tends to alternate in a strong – weak pattern within the category of normal full stress.

In utterances with a marked information structure the focused item is always perceived as very prominent, while pre-focal stresses are reduced in proportion to their distance from the focal accent. Post-focal reduction is not affected by the position of the focal accent but is more pronounced than pre-focal reduction – a phenomenon which has been noted in the literature as an absence of post-focal *accents*, that is, pitch-prominent syllables (Huss 1978, Nakatani and Aston 1978).

One might hypothesise that the results reported here were crucially influenced by the fact that the raters were not native speakers of English. For one thing, the failure to mark a consistent and clear difference in prominence level between the last stressed word in an utterance (the nucleus) and the first stressed word (the onset) could be attributed to the fact that in Danish neutral, context-free utterances all stressed syllables are normally produced and heard to be equally prominent. The partial reduction found in intermediate stresses refutes this argument, but the issue of whether there are systematic differences in the way native speakers of Danish and native speakers of English perceive prominence is still pertinent to the overall purpose of this study as well as to the specific problem concerning the nucleus. The test was therefore repeated with a group of native speakers of Southern British English, and this experiment – Test 2 – is presented in Chapter 5.

CHAPTER 5

Test 2 – English listeners

5.1 Introduction

The statistical methods used in this test are the same as for the Danish raters, so the reader is referred to Section 3.4.6 for a more detailed account of the procedures.

5.2 Subjects

As argued in Chapter 4 it might not be possible to generalise from the results of Test 1 with Danish raters to the perception of prominence by native speakers of English, especially with regard to properties which are known to differ between Danish and English, such as the presence of an obligatory nucleus.

The test was therefore repeated with a group of six native speakers of Southern British English. All six raters are professional phoneticians – one Professor Emeritus, one lecturer and four PhD students. The PhD students (all at Cambridge University, England) were paid for their participation.

5.3 Instructions to the raters

The aim of the test was to be able to compare the results directly with the scores from the Danish raters. It was therefore necessary that the instructions were as similar as possible so that the test could be said to be 'the same'. This not only required that a suitable translation of the Danish terms was found, but also that the instructions were not too closely connected with any one theoretical framework. It was explained in Section 3.4.3 that the terminology is fairly well established in Danish and that the academic terminology is in good agreement with the popular terminology and understanding of the phenomenon. Unfortunately the same assumptions cannot be made for English. The most common terms used are 'stress' and 'accent', and both have been used to cover the meaning needed for the experiment but also to cover different or more narrow meanings. Stress is often used to denote just one of the known attributes that result in perceived prominence, namely 'the force with which a sound or syllable is uttered' (Jones 1918: 247) and accent is often used about syllables where variation in pitch is the primary or even defining feature (especially within the autosegmental-metrical school). See Chapter 1 for an overview. I particularly wanted to prevent the raters from adhering to a theoretical framework where stress or accent is linked to specific events or positions in the utterance, such as the traditional British school of intonation where the terms 'primary

stress/accent, secondary stress/accent, tertiary stress' are defined by reference to function or position: primary stress = nucleus, secondary stress = onset (or other tonally marked accents), etc. The following simple set of instructions were found to be the best possible (if not perfect) solution:

" indicates extra strong stress.
' indicates (normal) full stress.
ˌ indicates reduced stress.

You can mark three degrees of stress, or fewer,
as you deem appropriate.

After some of the raters had completed the test, the following sentence was added for clarification:

(Completely unstressed words/syllables are not marked explicitly.)

The full set of instructions can be seen online at http://www.cphling.dk/pers/chrjen/stress/stress_uk.html and in Appendix A, Section A.3.

The phrase 'extra strong stress' is taken from the *Handbook of the International Phonetic Association* (Handbook 1999: 22) and was meant to indicate a level of stress used for contrast or other emphasis. Normal full stress was meant to cover the prominence of a syllable which is stressed but neither (contrastively) emphatic nor clearly reduced (in order to highlight a different syllable), and 'reduced stress' was preferred over 'secondary stress' to avoid associations to either the British school or to lexical stress in polysyllabic words, e.g. ˌfəʊnə'tɪʃən. In retrospect a more direct translation of the Danish instructions would have been preferable, but the terms 'extra strong stress' and 'reduced stress' were preferred since they are in more common usage in English language literature. The raters in this experiment were not asked to indicate boundaries since the Danish raters had indicated very few boundaries and with very little agreement.

5.4 Feedback from the raters

The paid raters in general found the test fairly manageable, but the two volunteer raters reacted rather differently from each other. While one of them found the test too long and rather tedious, the other thought it was 'Quite painless, really!'. As mentioned above, one rater had been uncertain about reduced stress versus unstressed, but it was confirmed in subsequent communication that she had performed the test according to the intentions of the instructions. One rater had omitted to mark stress on seven utterances. Since the data extraction process was automated and required that all fields contain an appropriate value, the gaps were filled

with the average scores of the other five raters. This affected 46 words out of the 907 words in the test.

5.5 Reliability

The reliability coefficients (see Section 3.4.6.1 for an explanation of the procedure) for the group of six English raters are as follows:

		Reliability
Group of raters	$R_{k(f)}$	0.966
Single rater	$R_{1(f)}$	0.826

Table 5.1. Reliability coefficient (Cronbach's alpha) for a group of raters and for a single rater based on six English raters.

Both values are quite high, which means that the data as a whole are reliable. The coefficients are somewhat lower than for the Danish raters, though (see Section 3.4.7), which might indicate either more disagreement among the raters or more uncertainty about how to perform the task. However, a statistical test of the reliability coefficients of the Danish and English raters shows that the difference between them is non-significant ($M = 1.20$, $df = 1$, $p > 0.1$).

5.6 Agreement

The distribution of scores in terms of number of 0, 1, 2 and 3 scores differs slightly from the Danish raters:

	Ratings (%)			
Spk.	0	1	2	3
r 1	39	2	44	16
r 2	39	17	38	6
r 3	46	12	29	13
r 4	31	17	37	15
r 5	41	13	34	12
r 6	32	15	39	14
Mean	38	12.7	36.83	12.7
S.d.	5.66	5.61	5.04	3.56

Table 5.2. Distribution of ratings as a percentage of the total number of ratings (907) for six native English raters. See Section 3.4.5 for an explanation of the values used to represent degrees of prominence.

The most noticeable difference between the Danish and the English raters in terms of distribution of scores is the number of 0-responses, that is, the number of times the raters heard a word as being completely unstressed. This difference is also clear

from the list of words which received an average score of less than one (that is, less than 'reduced stress') across all raters, in the two tests.

Danish raters		*English raters*			
Word	N	Word	N	Word	N
a/an	18	a/an	18	Bill	2
but	12	but	12	examined	1
did	6	did	6	Frank	3
from	12	from	12	import	8
in	9	in	9	kissed	15
is	8	is	9	Paul	3
know	6	know	12	sings	1
of	15	of	15	smelling	1
on	9	on	9	tenderly	2
struck	2	struck	14		
that	12	that	12		
the	118	the	118		
their	6	their	6		
was	42	was	42		
we	12	we	12		

Table 5.3. Words which received an average prominence rating of less than 1 (reduced stress). Based on ten Danish raters and six English raters. N = number of occurrences of each item.

Almost all the words which received an average prominence rating of less than 1 from the Danish raters are grammatical words. The only exceptions are 'know', which occurs in (immediate) post-focal position in two sentences (*imp_vb* and *imp_sb*) and 'struck'. All these words are also present in the corresponding ratings from the English raters, with the only difference being that 'know' and 'struck' received a score of less than 1 much more often. In addition, there is a list of lexical words which also received a low score from the English raters: 'Bill, examined, Frank, import, kissed', etc. Although the 'cut-off point' of 1 degree of prominence is partly arbitrary, this difference might indicate a systematic difference between Danish and English listeners in what constitutes 'reduced/weaker stress' and 'zero stress'.

It follows that when the English raters (implicitly) marked more words as being completely unstressed they used the other labels less frequently. This comes out in the average scores of 2 and 3, which are a little lower than in the Danish test, but the figure for score 1 is the same. This may have to do with the fact that four of the ten Danish raters did not use this label very often, while this only applied to one of the six English raters (*r* 1). It must be concluded that the English raters as a whole seem to be placed a little further towards the lower end of the scale than the Danish raters. In terms of individual variation two raters in particular seem to deviate somewhat from the norm: *r* 1, for using the label 'reduced stress' less frequently, and *r* 2, for

using the label 'extra strong stress' less frequently. Incidentally, these two raters are the most experienced of the six. Table 5.4 shows a distribution matrix of all the response pairs between any two raters.

Distribution of scores – English raters				
	y			
x	*0*	*1*	*2*	*3*
0	**33**	4	3	0
1	2	**5**	5	0
2	1	7	**25**	4
3	0	0	3	**9**

Table 5.4. Distribution matrix of all responses in the listening experiment as a percentage of the total number of comparison pairs: 15 rater pairs × 907 words = 13605 pairs.

This distribution is similar to the one for the Danish raters (see Table 3.5). Apart from the general shift downwards in prominence ratings, the only noticeable difference is in the number of disagreements where one rater scored 0 and the other rater scored either 1 or 2. This occurred in 4.1% of the pairs for the Danish raters, but in 10.6% for the English raters; a difference which can partly be attributed to the fact that some of the English raters – *r* 3 in particular – used the score 0 quite frequently. The distribution matrices for each individual rater[1] revealed that a substantial number of these disagreements involve this rater, but also rater *r* 5, for whom the disagreements go in both directions: in comparisons between *r* 5 and raters *r* 1, *r* 2 and *r* 3 there are quite a few cases where *r* 5 scored 0 and the other rater scored 2 *and* vice versa. This indicates genuine disagreement or uncertainty as opposed to a mere shift along the prominence scale.

Pairwise comparison and *T* coefficient

The same statistical measurements as in Test 1 were calculated: pairwise agreement between all 15 possible pairs of raters and total agreement where all six raters have to agree. The result is presented in Table 5.5, see Section 3.4.6 for an explanation of the procedures and Section 3.4.8 for the corresponding data for the Danish raters.

[1] Available on the webpage.

Pairwise comparisons	
Agreement (mean, %)	71.7
Total agreement	
S.d.	0.33
T	0.45
χ^2	188,688
N (no. of observations)	907
N_1 (no. of total agreements)	410
S.d. indicates mean value for the whole material	

Table 5.5. Agreement measurements for the responses of the six English listeners.

The pairwise agreement of 71.7% is lower than in the group of Danish raters (75.1%), while the total agreement across all raters, T = 0.45, seems to be higher (Danish: 0.41), reflecting the increase in number of total agreements. However, as explained in Section 3.4.6.2 it is not possible to compare two groups of unequal size, so if instead we compare with the six most experienced Danish raters (Groups 1 and 2) the corresponding figures are 77.2% pairwise agreement and a total agreement of T = 0.51. The scores for the English raters are quite similar to the scores for the two *lowest* scoring groups of Danish raters (Groups 2 and 3).

It must be concluded that the data material as a whole is reliable and that the raters seem to agree about the assignment of prominence levels. The agreement is a little lower than for the Danish group, which is confirmation that the Danish raters could perform the task in a completely reliable and satisfactory manner. I will now turn to the question of whether the general perception of prominence is similar in the two groups.

5.7 Prominence levels – English listeners

The utterances which were included in the mean scores in the tables and figures below are the same as for the Danish raters, that is, utterances with an *irregular* phrasal pattern have been excluded.

5.7.1 Context-free utterances

Abbrev.	*Sentence*
ps	$^{2.00}$Paul $^{1.97}$sings
bsa	$^{2.03}$Bill $^{0.87}$struck $^{2.30}$Ann
jkft	$^{2.00}$Jane $^{1.47}$kissed $^{2.03}$Frank $^{2.03}$tenderly
pc	^{0}The $^{2.04}$party ^{0}was $^{2.12}$cancelled

Abbrev.	*Sentence*
css	^{0}The $^{2.00}$cook ^{0}was $^{1.77}$smelling ^{0}the $^{2.00}$soup
sepc	$^{2.00}$Sheila $^{1.83}$examined ^{0}the $^{1.87}$patient $^{2.03}$carefully
tgios	^{0}The $^{2.00}$Germans' $^{1.70}$import ^{0}of $^{2.03}$sinks ^{0}from $^{2.03}$Denmark
gitsd	^{0}The $^{2.06}$Germans $^{1.78}$import ^{0}their $^{1.78}$sinks ^{0}from $^{2.17}$Denmark
pdp	$^{0.22}$Is $^{2.00}$Peter ^{0}a $^{1.67}$doctor ^{0}in $^{2.28}$Paris
dsi	$^{0.96}$Did $^{2.00}$Stalin $^{1.58}$insist ^{0}on ^{0}an $^{1.71}$equal $^{1.54}$distribution ^{0}of $^{2.08}$wealth

The grammatical words which were expected to be completely unstressed (almost) all achieved an average score of 0.00, which means that they can be left out of the following account. The two exceptions are the same as in Test 1 – 'Is' and 'Did' from *pdp* and *dsi* – and like those they are treated separately in Section 6.9.

The prominence levels of the stressed syllables in all the neutral utterances are presented diagrammatically in Figure 5.1.

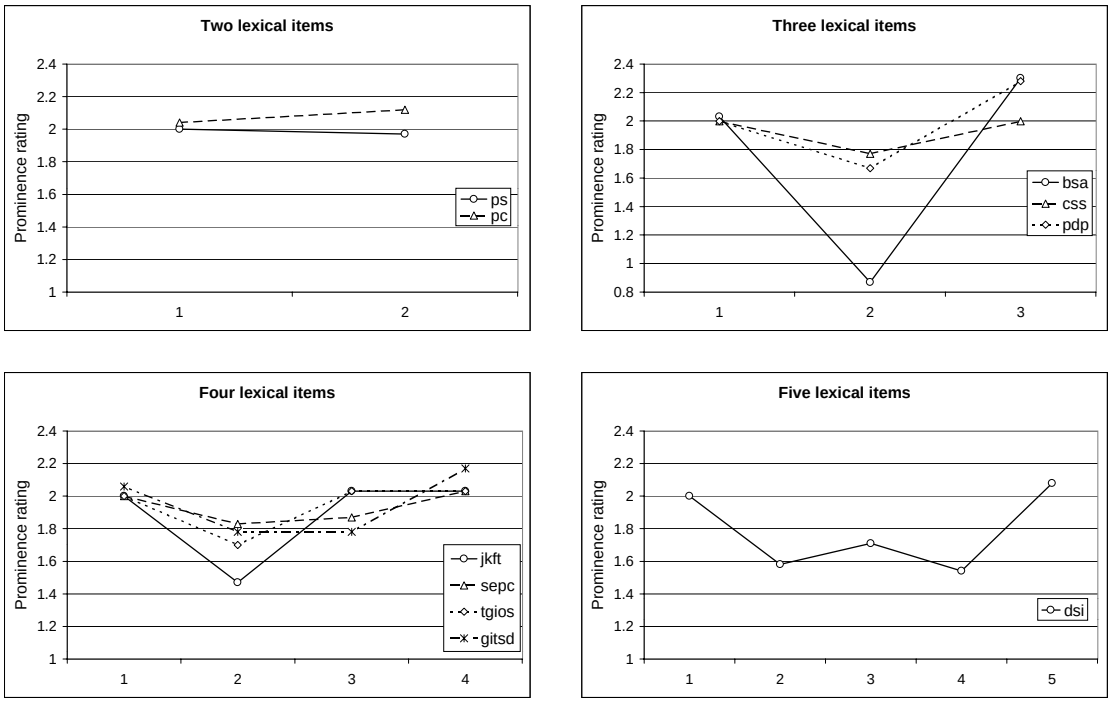


Figure 5.1. Stressed syllables in neutral sentences. Means across all speakers (excluding 3F) and all the English raters.

The prominence ratings presented in Figure 5.1 and in the list above do not differ essentially from the Danish ratings presented at the beginning of Section 4.2. One minor, or more specific difference concerns the word 'struck' in *bsa*, which received a much lower prominence rating by the English raters: 0.87 versus 1.70. This difference does not seem to be caused by a general higher sensitivity to the reduction of

intermediate stresses among the English listeners, because it is not found (as clearly) in the other sentences with three (or more) syllables, but it does seem as if the two groups of raters have responded to different properties of this word/sentence, and in the section on 'intervening items' below, one possible explanation is presented.

5.7.1.1 First and last lexical item

At the beginning of Section 5.2 it was mentioned that one might expect the English listeners to indicate a more consistent and clear difference between the first and last lexical item (corresponding to onset and nucleus), because the distinction between these two intonational events is relevant in English, whereas it is irrelevant in Danish. It is clear from Figure 5.1 that the variation in prominence ratings for the first (and partly the last) items is smaller in this test. All the first lexical items have an average prominence rating of very close to 2, and the inter-speaker variation is quite small. Many of the last items also received an average rating of close to 2 on the prominence scale, with just three exceptions: *bsa, pdp* and *gitsd*. In sentences *tgios* and *gitsd* the nuclei received a lower rating from the English raters than the Danish raters. The difference between the English and the Danish raters therefore seems to go in the opposite direction of what could be expected. Table 5.6 shows the mean values of, and differences between, the prominence ratings for first and last lexical item for the English raters, with an indication of whether the difference is significant. Compare this table with Table 4.2 which shows the corresponding data for the Danish raters.

Sent.	*First*	*Last*	*Diff.*	*N*	*p*
ps	2.00	1.97	0.03	30	0.745
bsa	2.03	2.30	0.27	30	**0.043**
jkft	2.00	2.03	0.03	30	0.573
pc	2.04	2.13	0.08	24	0.328
css	2.00	2.00	0.00	30	1.000
sepc	2.00	2.03	0.03	30	0.326
tgios	2.00	2.03	0.03	30	0.662
gitsd	2.06	2.17	0.11	18	0.430
pdp	2.00	2.28	0.28	18	0.096
dsi	2.00	2.08	0.08	24	0.328
All	2.01	2.09	0.08	264	**0.006**

p = two-tailed probability, paired t-test
N = number of pairs (first – last)
Significant values ($p < 0.05$) are in bold-face type.

Table 5.6. Prominence levels of the first and last lexical items in the context-free sentences – English raters. The difference 'last – first item' is listed in column four.

The difference between the first and last lexical item is only apparent in sentences *bsa* and *pdp*, and only statistically significant in the former, partly due to the smaller number of observations in this test. It therefore appears that the fact that the Danish listeners did not notice a general higher prominence level on the last items was not caused by the lack of relevance of such a distinction in Danish. The English listeners perceived the prominence level of this item as lower than did the Danish listeners and the difference between the first and last item as smaller.

5.7.1.2 Intervening lexical items

The general pattern of prominence relations between the stressed syllables in the neutral sentences is the same as for the Danish raters: there is a tendency for a strong – weak alternation, so that every other stressed item is reduced, or weaker. In sentences with three and five lexical items there is a significant difference between the intervening items and the first and last items, but the differences between the second, third and fourth items in sentence *dsi* are not significant. In sentences with four lexical items the third one is either deemed to be as weak as the preceding item (*sepc* and *gitsd*), and significantly weaker than the final item, or as strong as the following, final item (*jkft* and *tgios*), and therefore significantly stronger than the preceding item (in both cases $p < 0.05$, two-tailed t-test). This confirms the results from Test 1, which suggested that the penultimate item in utterances with an even number of lexical items is in a position of conflict between the preceding weak item and the following, final strong item.

One difference between the English and Danish raters concerns the magnitude of the reduction of the second lexical item. As mentioned earlier, the word 'struck' is an extreme case where the difference is almost 1 degree of prominence, but in sentences *jkft* and *gitsd* the difference is also noticeably larger. This is part of the general tendency towards larger reductions of lexical words among the English raters which was shown in Table 5.3, but there may also be a different, or complementary, explanation. Sentences *bsa* and *jkft* exhibit stress clash, and it may be that the English raters are more sensitive to this situation. Concerning the individual utterances and raters, the differences seem to be caused mainly by the fact that those Danish raters who deemed these words to be reduced from fully stressed heard them as having 'weaker stress' = 1, whereas the English raters heard 'no stress' = 0. There were also raters in both groups who judged these words to be fully stressed. In fact, the words 'struck' and 'kissed' in these utterances were among those which caused most disagreement, with standard deviations values (typically) between 0.8 and 1.1, so it seems that this type of reduction is subject to a large degree of inter-listener variability.

5.7.2 Utterances with marked information structure

The prominence ratings for the sentences with a marked information structure are presented below, averaged across all speakers (see comments at the beginning of Section 4.3 for exceptions).

Focus	*Sentence*
f1	$^{2.80}$Paul $^{1.14}$sings
f2	$^{1.05}$Paul $^{2.72}$sings
f1	$^{2.80}$Bill $^{0.67}$struck $^{1.47}$Ann
f2	$^{1.28}$Bill $^{2.89}$struck $^{1.53}$Ann
f3	$^{1.70}$Bill $^{0.60}$struck $^{2.77}$Ann
f1	$^{2.89}$Jane $^{0.56}$kissed $^{1.42}$Frank $^{1.53}$tenderly
f2	$^{1.29}$Jane $^{2.92}$kissed $^{0.75}$Frank $^{1.33}$tenderly
f3	$^{1.83}$Jane $^{0.79}$kissed $^{2.92}$Frank $^{0.92}$tenderly
f4	$^{1.72}$Jane $^{0.75}$kissed $^{1.55}$Frank $^{2.52}$tenderly
f1	^{0}The $^{2.80}$party $^{0.06}$was $^{1.17}$cancelled
f2	$^{0.03}$The $^{1.50}$party $^{0.03}$was $^{2.58}$cancelled
f1	^{0}The $^{2.75}$cook $^{0.06}$was $^{1.20}$smelling $^{0.03}$the $^{1.33}$soup
f2	$^{0.09}$The $^{1.53}$cook $^{0.09}$was $^{2.86}$smelling $^{0.03}$the $^{1.22}$soup
f3	^{0}The $^{1.83}$cook ^{0}was $^{1.53}$smelling ^{0}the $^{2.83}$soup
f1	$^{2.83}$Sheila $^{1.27}$examined $^{0.03}$the $^{1.40}$patient $^{1.43}$carefully
f2	$^{1.29}$Sheila $^{3.00}$examined ^{0}the $^{1.21}$patient $^{1.21}$carefully
f3	$^{1.58}$Sheila $^{1.29}$examined ^{0}the $^{2.92}$patient $^{1.38}$carefully
f4	$^{1.90}$Sheila $^{1.63}$examined ^{0}the $^{1.57}$patient $^{2.93}$carefully
f1	$^{0.07}$Is $^{2.73}$Peter ^{0}a $^{1.37}$doctor ^{0}in $^{1.60}$Paris
f1	$^{0.17}$Did $^{2.70}$Stalin $^{1.23}$insist ^{0}on ^{0}an $^{1.43}$equal $^{1.23}$distribution ^{0}of $^{1.80}$wealth

Again it is clear that the grammatical words can be excluded from further analysis. The prominence ratings of the lexical words are shown graphically in Figure 5.2.

Comparing the results for the English raters with the ratings from the Danish listeners in Figure 4.3 does not reveal any obvious systematic differences. The English listeners perceived a larger degree of reduction on the non-focal lexical items in some of the sentences, for example the words 'struck, kissed, Frank' in sentences *bsa, sepc* and *jkft*, respectively. But this parallels the difference found in the neutral utterances and therefore seems to be a question of relatively larger sensitivity to the strong – weak alternations. It is interesting that there is a fairly large difference between the perceived prominence of 'struck' in *bsa* and 'smelling' in *css*, and similarly between 'kissed' in *jkft* and 'examined' in *sepc*, whether in a neutral context or in a non-focal position in an utterance with a marked information structure. The most obvious explanation is that the strong – weak alternation is strongest when the stressed syllables are immediately adjacent, and partly blocked when there are intervening unstressed syllables. This means that the effect of stress clash (reduction of every other stressed item) can be observed even when the stressed words are backgrounded because of an explicit focus elsewhere in the sentence.

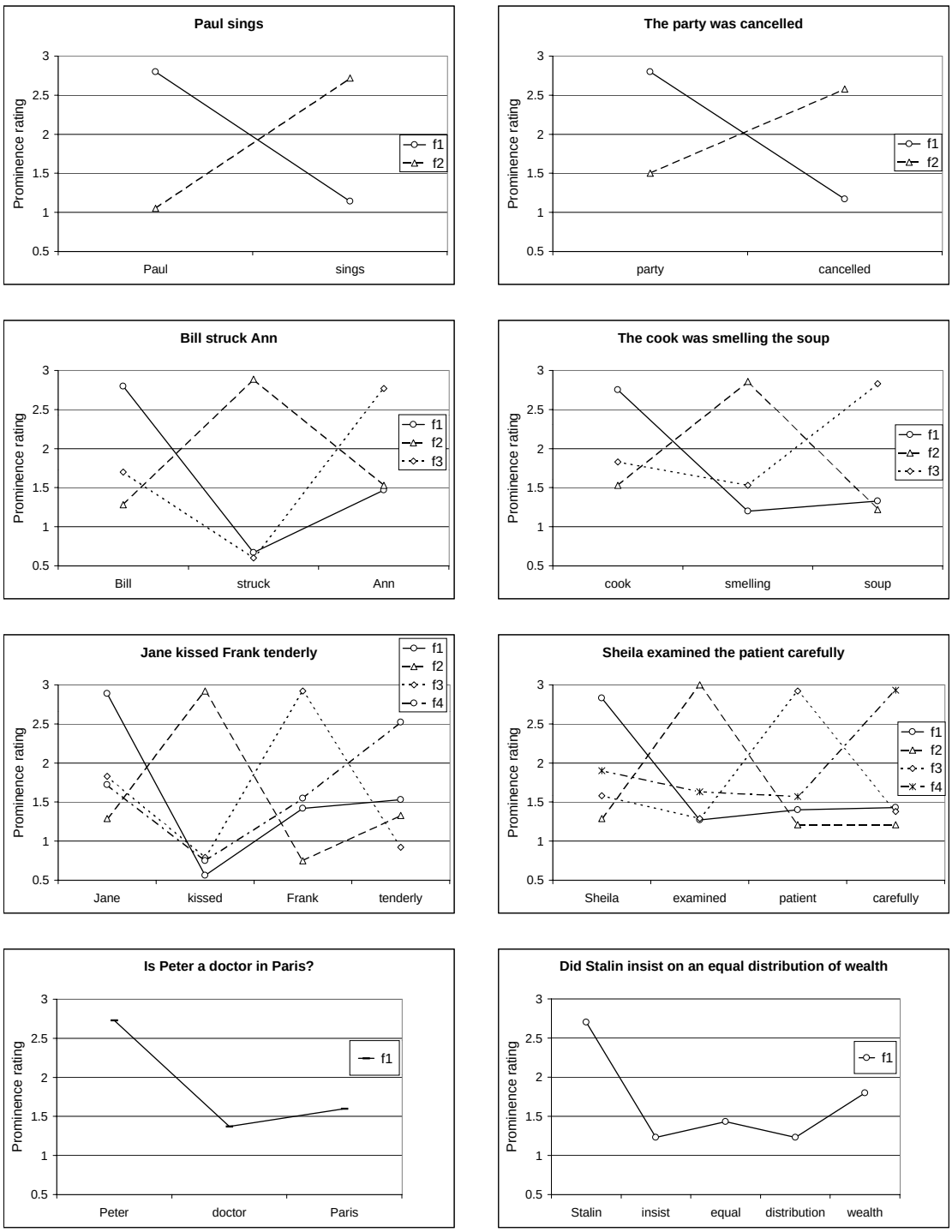


Figure 5.2. Prominence ratings of the stressed syllables in sentences with a marked information structure – English raters. The labels *f1, f2, f3, f4* indicate that the position of the intended focus is on the first, second, third and fourth stressed syllable respectively.

5.8 Summary of Tests 1 and 2

In Test 1 and Test 2 ten Danish and six English raters assigned prominence levels to the words in 183 short English utterances. The utterances represented 12 semantically different sentences, eight of which were present in both a 'neutral, context-free' version and in versions with contrastive focus on one of the constituents.

The raters indicated three degrees of stress (above unstressed), which were coded on a four-point numerical scale (including 0 for unstressed). The results for the Danish raters showed that the data were highly reliable and that there was good inter-rater agreement. Agreement was highest on completely unstressed words, almost all of which were grammatical words. There was also good agreement about words which were assigned the highest level of stress – labelled 'strong stress' – but noticeable disagreement about the labels 'normal stress' and especially 'weaker stress'. It was argued that this might be an indication that weaker, or reduced, stress is a linguistically (phonologically) less pertinent category and therefore cognitively less stable.

There were three subgroups of (three) Danish raters with different degrees and types of linguistic training. There was some variation in intra-group agreement, which may be correlated with level of experience with this type of task, but there did not seem to be much systematic variation in the distribution of responses between the groups: the individual variation within each group was much larger.

Prominence ratings were presented for both neutral, context-free utterances and utterances with a marked information structure (semantic or contrastive focus). In the neutral utterances it seemed clear that the two peripheral stressed words (the first and last lexical items) stood out from the intervening items as (slightly) more prominent. This corresponds with the special status these positions have in the traditional British school of intonation as onset and nucleus, respectively. They both achieved an average prominence rating of very close to (or slightly exceeding) 2, that is, normal stress in my experiment. There was very little difference between the prominence ratings of the first and last lexical item, but the last could be shown to be slightly more prominent in some cases. The hypothesis that the difference would be clearer in Test 2 with English raters (due to differences in linguistic structure between Danish and English) could not be supported. In Test 2 the difference between first and last item was even smaller and almost completely absent. There were, however, clear examples where the nucleus was more prominent, but they were infrequent and are thus best seen as only one of several realisational strategies available to the speaker.

Intervening stressed syllables were somewhat reduced/weaker compared with the peripheral stresses, although typically much less than by one full degree of prominence. This reduction was more pronounced in Test 2 with English raters, especially in utterances where a stressed syllable was followed immediately by another stressed syllable. This points to a greater sensitivity among English raters to the strong – weak alternations which could be observed in both tests, especially when the utterances exhibit stress clash. The strong – weak alternations are similar

to what can be found in English polysyllabic words, only to a lesser degree. In most of the cases the reductions could not be considered full categorical deaccentuation.

In utterances with a marked information structure it was observed that (1) the word which is focused is always the most prominent, with an average rating of 3 (strong stress) or just under in most cases, and (2) the remaining lexical items have reduced prominence compared with the neutral utterances. This reduction seems to be in inverse proportion to the distance of the stressed word from the word with focal stress, that is, the immediate neighbours are most clearly reduced.

The prominence relation between the non-focal lexical items is similar to the relation between the same words in the neutral version of a given utterance, including the tendency towards a strong – weak alternation. It is therefore difficult to determine whether an observed pattern is caused by the principle of reduction in inverse proportion to the distance from the position of focus or by the strong – weak alternation; or quite possibly to a combination of the two. It is clear that marking contrastive focus involves both local signals on the focused word and more general, or global, signals affecting the surrounding stressed words.

Because of the similarity between the responses of the Danish and English raters their responses can be pooled for further analyses, either of the acoustic correlates of the perceived prominence or for an investigation of the connection between prominence and information structure (see Chapter 7).

CHAPTER 6

Test 3 – British school of intonation analysis

6.1 Introduction

In the analysis of prominence relations in Chapters 4 and 5 several references were made to the 'British school of intonation' and how my findings correspond with the most common account of, or assumptions about, prominence in this tradition. In particular, it was examined whether the stress/accent hierarchy, with the nucleus as primary accent, the onset as secondary accent, and other stresses as tertiary stresses was adequately reflected in the prominence ratings given by the ten Danish and six English listeners. Although the data to a certain extent fitted this hierarchy, this was by no means unproblematic, and in no way conclusive evidence that the assumptions about, or definition of, stress levels in the British school are appropriate. However, in order to make a reasonable, valid analysis empirical data is needed to show how this system would actually be applied to the present material. Therefore a group of native English phoneticians were asked to assign 'tonetic stress marks', to the same test material that was used in Test 1 and Test 2. The only difference between this test and the two previous tests is in the instructions given to the raters.

6.2 Stress/accent levels

As it appears elsewhere in this investigation, I have used Alan Cruttenden's book *Intonation* from 1997 as a representative of the British school of intonation analysis. It is one of the most recent works on intonation in this tradition, and it has a good description of stress/accent levels, which is in line with that in Gimson's standard textbook on English phonetics, *An introduction to the pronunciation of English* (Gimson 1989)[1]. In both these works a system of four degrees of stress/accent is outlined, with slightly varying but compatible definitions. The reader is referred to Figure 1.2, reproduced from Cruttenden (1997) and the presentation of the stress/accent hierarchy in that section for an outline of the system. The definitions in Figure 1.2 were also used in the instructions to the raters.

The intonational system underlying this prominence scale provides quite clear predictions about which degree of stress/accent one will find on any given word in an

[1] This is the fourth edition of the book, edited by Susan Ramsaran, before Alan Cruttenden took over as editor.

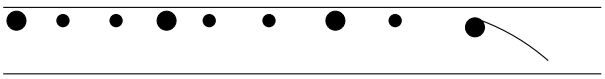
utterance (at least in short neutral utterances), since especially the two strongest degrees of accent are associated with specific events or positions in the utterance. The primary accent, or nucleus, is normally found on the last lexical word in the utterance (= the last word which can normally be stressed), and it is only possible to have more than one nucleus if the utterance contains a boundary, so that it is divided into two or more intonation units (or in the special case of a *compound tune*, see later this chapter). Secondary accents are typically found on the first lexical word in the utterance, in which case they are sometimes called onsets, but can also occur elsewhere in an utterance. The stressed syllables between onset and nucleus, which I referred to as intervening stressed words in the previous experiment, will normally receive tertiary stress or, in the case of a stepping pattern, secondary accent. This means that it is possible to make the following predictions about the degrees of stress in the neutral version of the sentence *Sheila examined the patient carefully* from my material. In the following examples these conventions are used:

1 = primary stress/accent
2 = secondary stress/accent
3 = tertiary stress
(unstressed not marked)

Please note that these conventions are different from the system used to code stress levels in my analyses, where '3' denotes the highest level of prominence.

There are two likely possibilities:

(1) ^{2}Sheila e^{3}xamined the ^{3}patient ^{1}carefully



(2) ^{2}Sheila e^{2}xamined the ^{2}patient ^{1}carefully

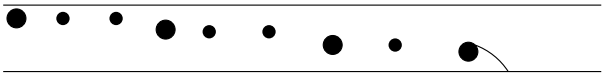


Figure 6.1. Two possible, and typical, realisations of the sentence 'Sheila examined the patient carefully' as predicted by the British school of intonation.

In a context-free utterance the nucleus is expected to fall on the last lexical item ('carefully'); I disregard for a moment the choice of nuclear tone (type of pitch movement associated with the nucleus), since it does not affect the degree of stress/accent (although it might affect the degree of perceived prominence). The *head*, that is, the pre-nuclear pattern, can either be level (or slightly falling) as in (1) or have a sequence of steps down as in (2). In the first case that would signal secondary

stress/accent on the first word, because a tone level significantly above the neutral baseline gives pitch prominence to the first stress, and tertiary stress on the other two lexical words. In the second case the steps down in pitch lend pitch prominence to each lexical word producing three consecutive secondary accents (see Cruttenden 1997: 54). While the theory makes strong predictions about the stress levels of the first and last stressed items, there are two likely possibilities for the intervening items. The test will therefore be able to supply information about which of these structures are perceived by the listeners even if they are not asked to indicate it directly, provided that they adhere to the principles of the system. The acoustic description in Section 2.5.2 showed that some degree of downstep was very common in the material, but it is uncertain (and unspecified in the theory) how large the F_0 difference between two successive stressed syllables must be before it constitutes downstep (proper), and the perception of a secondary accent, rather than just some degree of downtrend or declination which does not trigger the perception of pitch prominence.

Neither Cruttenden nor Gimson make any claims, nor indeed imply, that their stress scale corresponds with a scale such as the one used in Tests 1 and 2. One might especially question whether the label '(extra) strong prominence' corresponds well with 'primary stress/accent' or whether this label does not imply some sort of emphasis, even in the traditional British system. But no provision is made in this system for an emphatic accent, and differences between neutral, context-free utterances and utterances with a specific *narrow* focus are often treated as a difference in nucleus *placement* rather than a difference in degree of emphasis/stress (Cruttenden 1997: 73 ff.). And since the Cruttenden/Gimson system does in fact use four levels of stress it is very relevant to see how these levels match my more simplistically defined levels of prominence, and how well they can reveal differences between the various types of sentence in my investigation.

6.2.1 Aims of Test 3

Some of the aims of the experiment have already mentioned, but the three main purposes are summarised below:

(1) To examine the notational system of the traditional British system of intonation analysis on its own premises. What information does it provide about the prominence levels in the utterances and, to some extent, about the intonational structure of these.

(2) Is the representation of stress/accent levels produced by this system an adequate basis for examination of perceived prominence levels, their acoustic manifestation, and relation to information structure?

(3) How do the results of this test compare with the results from Test 1 and Test 2? Are the prominence ratings implied by the British system in agreement with those from Tests 1 and 2, which were less bound by a particular model of intonation?

6.3 Subjects and instructions

There are four raters in this experiment. All are native speakers of British English and professional phoneticians with long experience in this type of task. The instructions to the raters were to mark stress and accent in a series of short English utterances according to the principles in Cruttenden (1997: 18,44). The raters were asked to assign numbers to the stressed syllables using the same conventions as in the example above. Alternatively, they were allowed to use a more conventional 'tonetic stress-mark' system, but were advised not to devote too much attention to the question of type of nuclear tone. Since there is a strong formal connection between the occurrence of primary accents and boundaries in this system (one intonation-unit, one primary accent), the raters were asked to indicate boundaries with the symbol /. An example sentence from Cruttenden (1997) was included as an illustration of the system. See http://www.cphling.dk/pers/chrjen/stress/stress_en.html or Appendix A, Section A.3 for a full set of instructions.

6.4 Feedback from the raters

None of the raters reported any problems performing the task, that is, understanding what they were asked to do, but there were some interesting comments from two of the raters regarding assignment of stress/accent level in some contexts. Rater *r* 2 made several comments, and pointed to some of the more problematic issues:

> One is the familiar problem of Fall-Rise with tail versus Fall plus Rise. The decision one makes about the identity of the nuclear tone has implications for the last stress in the sentence, which is 3 if part of the tail but 1 if bearing the nucleus of a Rise.
> (Rater *r* 2, written comment)

The 'familiar problem' which rater *r* 2 refers to involves intonation units which have falling pitch (on a nuclear syllable) followed later in the unit by rising pitch on a syllable with accent potential (bearing lexical stress). These tonal patterns can be interpreted either as a single (falling-rising) nuclear tone distributed over several words or a combination of two nuclear tones – the first one falling and the second one rising. The issue is treated in O'Connor and Arnold (1973: 28 ff.), where it is noted that phonetic cues may in some cases help to disambiguate the two tunes. But even if this is not the case O'Connor and Arnold argue that the two types should be distinguished in notation 'because of the different attitudes that the two tunes convey' (1973: 30). The main point is that according to the above comment, and also acknowledged in O'Connor and Arnold (1973), it is very difficult to auditorily perceive a difference, the consequence of which is to assign either the highest degree of stress/accent or the lowest degree above unstressed to a particular word. In Tests 1 and 2 such a difficulty was only found in the unusual case of 'high pre-head' and it points to a flaw in this system of intonation analysis that such a substantial difference in prominence assignment can be caused by very subtle phonetic cues (rhythmi-

cal or tonal) or even in some cases by syntactic or semantic factors alone (Cruttenden 1997: 36).

Another point which was raised by rater *r* 2 is that 'the decision 2 versus 3 in prenuclear position is often difficult'. This decision was, of course, the largest source of disagreement in Test 1 and Test 2.

6.5 Data

The prominence ratings in this test were coded in the following manner:

primary stress/accent	= 3
secondary stress/accent	= 2
tertiary stress	= 1
unstressed	= 0

This system is as similar as possible to the one used in Tests 1 and 2 by having '3' denote the highest level of prominence and '0' as the lowest level of prominence (completely unstressed). This scale is equivalent to the one used in the first two tests only by having the same number of levels. The definition of each level is different between the two systems and they can therefore not *a priori* be expected to represent the same perception of prominence level. Rather, they should be treated as separate systems where differences or similarities will be represented in the numerical value assigned to each word.

6.6 Reliability

The reliability of the ratings was determined in the same manner as in Tests 1 and 2 by calculating Cronbach's alpha for a group of raters and for a single (typical) rater (see Section 3.4.6.1 for an explanation of the procedure).

		Reliability
Group of raters	$R_{k(f)}$	0.976
Single rater	$R_{1(f)}$	0.911

Table 6.1. Reliability coefficient (Cronbach's alpha) for a group of raters and for a single rater based on four English raters. Prominence ratings according to the British school of intonation analysis.

The reliability coefficients for the prominence ratings are comparable to those obtained in Tests 1 and 2, and the differences between them are non-significant (M = 1.34, df = 2, p > 0.1). The coefficient for the whole group of raters is (again) close to 1, and the coefficient for a single rater is the highest obtained in the three tests. The data can therefore be considered highly reliable with a high degree of covariation in the ratings of the four listeners (that is, good inter-rater correlation).

6.7 Agreement

The overall distribution of scores (share of 0, 1, 2 and 3 scores) is quite different from Tests 1 and 2:

	Ratings (%)			
Spk.	0	1	2	3
r 1	37	21	18	25
r 2	33	20	19	28
r 3	31	21	20	28
r 4	33	25	20	23
Mean	33.5	21.75	19.25	26
S.d.	2.52	2.22	0.96	2.45

Table 6.2. Distribution of ratings as a percentage of the total number of ratings (907) for four raters (British school of intonation).

The number of words which were perceived as completely unstressed is similar to what was found in the first two tests; slightly more than in Test 1 but fewer than in Test 2. An analysis shows that it is (again) the case that all grammatical words are perceived as unstressed, while only a few lexical words are perceived as unstressed. Slightly fewer lexical words were perceived as unstressed in this test than in Test 2, and I have no explanation for this difference in perception between two groups of native English listeners. There does not seem to be anything in the theoretical framework which dictates such a difference, but one of the raters in the current experiment – *r* 2 – did comment: '[...] I am reluctant to recognise lexical monosyllables as entirely unstressed'. Whether this attitude is more common within the traditional British framework than elsewhere I do not know.

The number of 2-responses is considerably lower in the present experiment than in Test 1 and Test 2; only 19% of the responses belong to this category compared with 37% and 40% in the other tests. Instead, the number of 3-responses is about twice as large as in the other tests which is to be expected because of the requirement in the British tradition that all utterances must have a nucleus (= 3), even if there is no explicit focus. The number of 1-responses is also larger, which is most likely due to the restriction in this system that secondary accents (2) cannot occur in post-nuclear position, where tertiary stress (1) is instead expected.

Besides the different distribution of responses in this experiment one may also notice that the variance is somewhat smaller; from just under 1 to just under 3 standard deviation units. In Test 1 (Danish raters) values were between 2.63 and 9.12 and in Test 2 (English raters) the standard deviations varied between 3.56 and 5.66. In other words, the four raters agree better about the number of each type of response than in the other experiments. The distribution matrix in Table 6.3 shows that there

is also better agreement about the *distribution* of responses in this experiment, although all categories give rise to disagreements of up to two scale points.

Distribution of scores – British school					
			y		
x	*0*	*1*	*2*	*3*	*Tot*
0	**31**	2	1	0	34
1	1	**16**	3	1	21
2	0	4	**13**	2	19
3	0	1	3	**23**	27
Tot	32	23	20	26	101%

Table 6.3. Distribution matrix of all responses in the listening experiment (traditional British system) as a percentage of the total number of comparison pairs: 6 rater pairs × 907 words = 5442 pairs.

The initial observation regarding agreement made from Table 6.3 can be expressed more clearly by using the indices of agreement which were also used in Tests 1 and 2, namely pairwise agreements between all possible rater pairs and exact agreement among all raters.

Pairwise comparisons	
Agreement (mean, %)	83.2
Total agreement	
S.d.	0.18
T	0.72
χ^2	29,393
N (no. of observations)	907
N_1 (no. of total agreements)	655
S.d. indicates mean value for the whole material	

Table 6.4. Agreement measurements for the responses of the English listeners using the traditional British school of intonation system.

The pairwise agreement between the four raters is very high; over 83%, which is considerably higher than the numbers for the Danish raters (75.1%) and English raters (71.7%) in Tests 1 and 2. It is higher than for the best 'natural' group of three raters in Test 1 (general phoneticians, see Section 3.4.8.3), and only two groups of four raters out of the possible 210 groups in that experiment can match this level of agreement. It can be concluded that the four raters in this experiment exhibit a high level of agreement compared with the raters in Test 1 and Test 2 and with raters in similar experiments (Heldner 2001a, Silverman *et al.* 1992).

6.8 Prominence levels – British school of intonation

As in Test 1 and Test 2 certain utterances were excluded from the mean prominence scores presented below, namely those with an irregular phrasal pattern (that is, the few clearly polyphrasal utterances in the material).

6.8.1 Context-free utterances

Abbrev.	*Sentence*
ps	$^{2.20}$Paul $^{3.00}$sings
bsa	$^{2.25}$Bill $^{1.05}$struck $^{3.00}$Ann
jkft	$^{2.00}$Jane $^{1.60}$kissed $^{2.00}$Frank $^{3.00}$tenderly
pc	^{0}The $^{2.40}$party ^{0}was $^{3.00}$cancelled
css	^{0}The $^{2.45}$cook ^{0}was $^{1.95}$smelling ^{0}the $^{3.00}$soup
sepc	$^{2.05}$Sheila $^{1.75}$examined ^{0}the $^{1.55}$patient $^{3.00}$carefully
tgios	^{0}The $^{2.00}$Germans' $^{1.60}$import ^{0}of $^{2.00}$sinks ^{0}from $^{3.00}$Denmark
gitsd	^{0}The $^{2.25}$Germans $^{1.65}$import ^{0}their $^{2.20}$sinks ^{0}from $^{2.90}$Denmark
pdp	$^{0.50}$Is $^{2.17}$Peter ^{0}a $^{1.67}$doctor ^{0}in $^{2.83}$Paris
dsi	$^{1.06}$Did $^{1.81}$Stalin $^{1.94}$insist ^{0}on ^{0}an $^{1.62}$equal $^{1.44}$distribution ^{0}of $^{3.00}$wealth

As in Test 1 and Test 2 the grammatical words are all perceived as completely unstressed, except for the words 'Is' and 'did' in sentences *pdp* and *dsi*, which are commented on in Section 6.9 but otherwise disregarded. Leaving out the grammatical words the results are presented diagrammatically in Figure 6.2.

6.8.1.1 First and last lexical item versus onset and nucleus

In all sentences the last lexical item has achieved an average rating of 3, or very close to 3, indicating that it has been perceived as a primary accent (or nucleus), the highest possible level of stress within this system. This is as predicted from the theoretical descriptions of intonation within the British framework (Cruttenden 1997, Gimson 1989), but it is very different from the prominence ratings obtained in Test 1 and Test 2 (see Figure 4.1 and Figure 5.1), where the same words receive average ratings of 2 or just over 2 (normal, full stress). The first lexical item of each sentence, which is expected to be the onset of the intonation unit achieved ratings of 2 to 2.5 degrees of stress/accent in most cases. This is slightly higher than in Tests 1 and 2, and the cases where the rating is (significantly) above 2 indicate that one or two raters have perceived these words as carrying primary accents, usually followed by a boundary. It is interesting that these short utterances have often been perceived as bi-phrasal with a nucleus on the first lexical item of the utterance. Keep in mind that the utterances where a division into several phrases was very obvious have been excluded from analysis, so all the utterances represented in Figure 6.2 can be regarded as monophrasal. This view is confirmed by the results from Test 1: the Danish raters did not perceive a

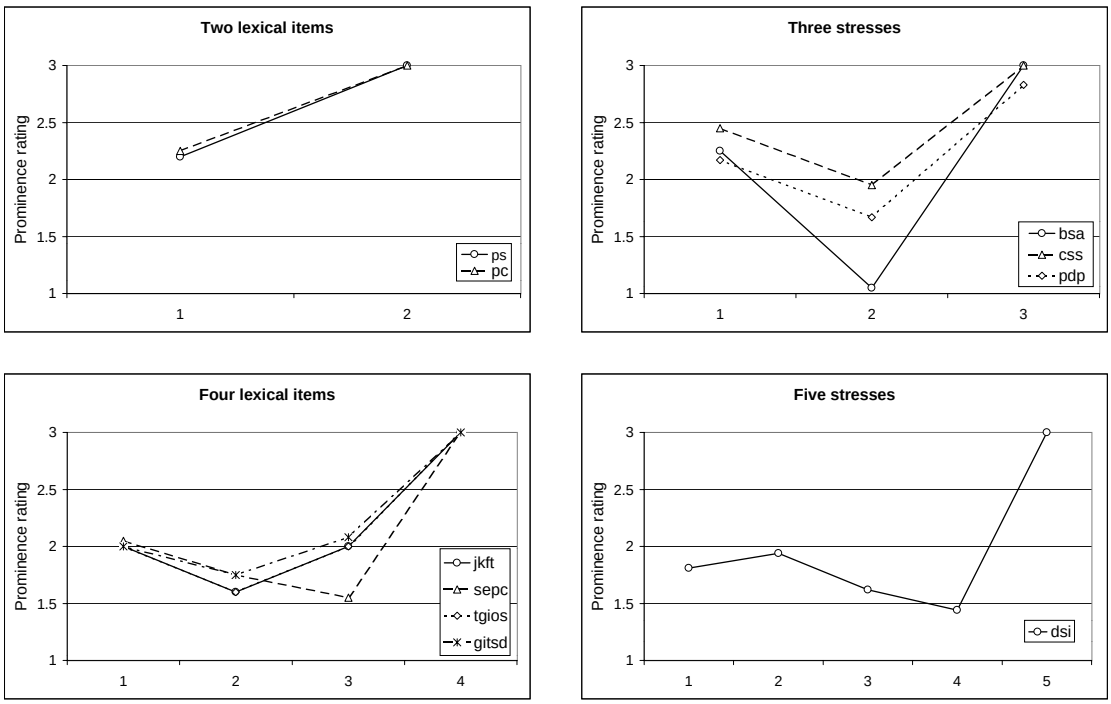


Figure 6.2. British school of intonation. Stressed syllables in neutral sentences. Means across all speakers (excluding 3F) and all four raters. Note that values for *jkft* and *tgios* are identical; the curves for these two sentences are therefore completely congruent.

boundary in these utterances, while some raters did perceive a boundary in the utterances which were excluded. Secondly, there is not total agreement among the raters in this experiment about the presence of a boundary in these utterances, although there does seem to be a pattern both with regard to speakers and raters. Many of the utterances which were perceived as bi-phrasal by some raters in this experiment were from speaker 4M, who also produced two of the three excluded utterances. Among these were the two shortest utterances with only two lexical items, 'Paul sings' and 'The party was cancelled'. None of the Danish raters perceived a boundary in any of these utterances, but there is a clear fall in F_0 on both stressed syllables which may have prompted the perception of two nuclei, and hence, by definition, a boundary between them. The raters were rarely in total agreement, but there was often agreement between raters $r2$ and $r3$, and sometimes $r1$, while rater $r4$ rarely perceived two nuclei plus a boundary.

The above observations may point to fundamentally different perceptions of what constitutes a nucleus and a boundary, partly between raters within this particular framework, but also across different intonational systems and/or across different first language backgrounds. The difference between the Danish raters in Test 1 and the English raters in Test 3 could be explained in more than one way. One explanation might be that since local syllable internal F_0 movements do not signal phrase

boundaries in Danish, the Danish raters are simply not attuned to this type of signal in English. They were therefore unable to pick up on the more subtle boundary markers which some of the native English listeners could perceive. The force of this argument is diminished by the fact that at least five or six of the ten raters were highly proficient speakers of British English, many with long training and experience in British English prosody. It is more likely that the Danish raters did perceive the tonal variations, but that these were not felt to constitute a clear boundary signal. The other explanation would focus on why the English raters then perceive these boundaries. As mentioned in Section 4.2, the definition of the nucleus and the definition of intonation units are in some ways interdependent in most descriptions, including two I have been referring to on several occasions, namely Cruttenden (1997) and O'Connor and Arnold (1973). In these descriptions an intonation unit (or tone group) contains one nucleus (with the exception of the *compound tune*, which contains a sequence of a 'High Fall followed by a Low Rise' (O'Connor and Arnold 1973: 28). The nucleus is often described, or defined, as 'the pitch accent which stands out as the most prominent in an intonation-group' (Cruttenden 1997: 42), and this prominence is typically associated with a clear change in pitch on or around the nuclear syllable. Pitch accents, or stressed syllables, other than the nucleus usually do not involve a pitch change on the syllable itself (but see Cruttenden 1997 :54 for an exception), which means that such a pitch change becomes an almost defining quality of a nucleus. It is therefore not surprising that a sequence of two syllables with pitch change is perceived by some raters as two nuclei divided by a boundary, as in the example 'Paul sings', even in the absence of clear rhythmical evidence for a boundary. Particularly if the pitch change is bi-directional as in the example in Figure 6.3.

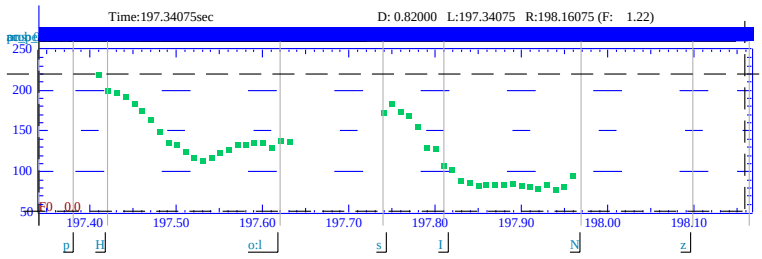


Figure 6.3. Pitch contour for the 'neutral' version of the sentence 'Paul sings' by speaker 4M.

In the light of this discrepancy between the Danish raters in Test 1 and the English raters (British school) in Test 3, it is perhaps regrettable that the English raters in Test 2 were not asked to indicate boundaries. Their responses could have provided an answer as to whether the discrepancy is caused by differences in the task – marking prominence levels in a simple, atheoretical framework, or marking stress/accent types according to the traditional British system – or by the linguistic background of the raters.

The different perception of phrasing in Test 3 also has consequences for the observations about the prominence levels of onsets and nuclei. In Tests 1 and 2 it was concluded that onsets and nuclei were perceived as almost equally prominent, but this was based on the assumption that all utterances were monophrasal and that the onset was equal to the first stressed syllable of the utterance and the nucleus was equal to the last stressed syllable. This assumption is not confirmed in all cases here, and in order to make observations about the prominence levels of onsets and nuclei *as they have been identified explicitly within the British framework* it is necessary to include only utterances that meet certain requirements. Since the four raters were rarely in complete agreement I decided that it was necessary for at least three raters to have identified the same onset and the same nucleus within one utterance in order for it to be included in the analysis. Nuclei were marked explicitly in the experiment while onsets were taken to be the first stressed syllables in an utterance, or following a boundary, which was marked as carrying a secondary accent; that is, the first secondary accent of an intonation unit.

Following these criteria, 41 intonation units in 40 utterances were selected. Two of these had to be excluded because the item identified formally as the onset is better analysed as a high pre-head – 'Is' in *pdp* and 'Did' in *dsi*, see Section 6.9. There was very large disagreement about these utterances in all experiments (Tests 1, 2 and 3), and they are best treated separately. That leaves 39 intonation units for analysis. Reporting the ratings from the current experiment based on the British framework is uninteresting, since the prominence levels of onsets and nuclei are *defined* and not negotiable. Instead, the prominence levels as they were perceived by the Danish raters in Test 1 and the English raters in Test 2 can be applied to the onsets and nuclei identified in this experiment (Test 3). So what follows is an account of prominence levels as *perceived by the raters in Tests 1 and 2* in onset and nucleus syllables as they were *identified by the raters in Test 3 (British tradition)*. Average prominence ratings for all 39 utterances are presented in Table 6.5.

Prominence ratings, onset and nucleus								
	Word		*Danish raters (Test 1)*			*English raters (Test 2)*		
Utt.	*Onset*	*Nucleus*	*Ons*	*Nuc*	*Diff.*	*Ons*	*Nuc*	*Diff.*
124	Paul	sings	1.90	2.20	0.30	2.00	2.33	0.33
65	Paul	sings	2.10	2.00	-0.10	2.00	1.67	-0.33
49	Paul	sings	2.10	2.00	-0.10	2.00	1.83	-0.17
178	Paul	sings	2.10	2.00	-0.10	2.00	2.17	0.17
183	party	cancelled	2.00	2.20	0.20	2.00	2.17	0.17
131	party	cancelled	2.10	2.00	-0.10	2.00	2.00	0.00
141	Bill	Ann	2.00	2.00	0.00	2.00	2.17	0.17
72	Bill	Ann	2.00	2.60	**0.60**	2.00	2.83	**0.83**
52	Bill	Ann	2.10	2.00	-0.10	2.00	1.83	-0.17
1	Bill	Ann	2.10	2.00	-0.10	2.00	2.00	0.00
92	Bill	Ann	2.30	2.10	-0.20	2.17	2.00	-0.17

Prominence ratings, onset and nucleus								
	Word		*Danish raters (Test 1)*			*English raters (Test 2)*		
Utt.	*Onset*	*Nucleus*	*Ons*	*Nuc*	*Diff.*	*Ons*	*Nuc*	*Diff.*
162	cook	soup	1.90	2.10	0.20	1.83	2.00	0.17
125	cook	soup	2.10	2.00	-0.10	2.00	2.00	0.00
84	smelling	soup	1.90	2.10	0.20	2.00	2.00	0.00
100	Jane	Frank	2.00	2.20	0.20	2.00	2.33	0.33
154	Jane	tenderly	2.00	2.10	0.10	2.00	2.00	0.00
168	Jane	tenderly	2.00	2.10	0.10	2.00	2.00	0.00
83	Jane	tenderly	2.00	2.20	0.20	2.00	2.33	0.33
158	Jane	tenderly	2.10	2.00	-0.10	2.00	2.00	0.00
43	Sheila	carefully	1.90	2.10	0.20	2.00	2.00	0.00
167	Sheila	carefully	2.00	1.90	-0.10	2.00	2.00	0.00
30	Sheila	carefully	2.00	2.00	0.00	2.00	2.00	0.00
148	Sheila	carefully	2.10	2.00	-0.10	2.00	2.00	0.00
132	Sheila	carefully	2.10	2.50	**0.40**	2.00	2.17	0.17
155	examined	carefully	2.00	2.00	0.00	1.83	2.00	0.17
21	Germans	Denmark	1.90	2.70	**0.80**	2.17	2.67	0.50
61	Germans	Denmark	2.00	2.10	0.10	2.00	1.83	-0.17
87	Germans	sinks	2.00	1.90	-0.10	2.00	2.00	0.00
4	import	Denmark	1.50	2.30	**0.80**	1.33	2.00	0.67
98	Germans'	Denmark	2.00	2.20	0.20	2.00	2.00	0.00
51	Germans'	Denmark	2.00	2.50	**0.50**	1.83	2.17	0.34
166	Germans'	Denmark	2.10	2.10	0.00	2.00	2.00	0.00
181	Germans'	Denmark	2.10	2.10	0.00	2.17	2.00	-0.17
15	Germans'	Denmark	2.20	2.00	-0.20	2.00	2.00	0.00
57	Germans'	Denmark	2.20	2.30	0.10	2.33	1.83	-0.50
47	Peter	Paris	2.00	2.20	0.20	2.33	2.17	-0.16
54	Stalin	wealth	2.00	2.00	0.00	1.83	2.00	0.17
9	Stalin	wealth	2.00	2.50	**0.50**	2.17	2.33	0.16
112	equal	wealth	2.00	2.10	0.10	2.00	2.00	0.00
Mean values			2.02	2.14	**0.12**	2.00	2.07	**0.07**

Table 6.5. Prominence ratings for onsets and nuclei in 39 intonation units in 'neutral' utterances. Scores from the Danish raters are from Test 1 and scores from the English raters are from Test 2. The numbers in the 'utterance' column refer to the number/position in the online test (http://www.cphling.dk/pers/chrjen/stress/stress_en.html). Bold-face type indicates significant differences ($p < 0.05$, two-tailed t-test).

The prominence ratings of onsets and nuclei as they were identified by raters within the traditional British framework do not differ much from the ratings which were based on the first and last lexical (stressed) items in utterances which were not very clearly polyphrasal. This partly reflects the fact that most of these short utterances were perceived as monophrasal in all three tests, and the figures in Table 6.5 above are therefore to a large extent based on the same utterances as in Test 1 and Test 2. Thirty-five of the 39 selected intonation units (with onset – nucleus pairs) correspond to utterances also analysed in Tests 1 and 2, while the remaining four are new additions where the onset is not the first lexical item of the utterance.

The utterances where the first lexical item was heard as a nucleus in Test 3 have not had a significant effect on the comparison of first and last lexical item in Tests 1

and 2. In fact, the pattern in terms of distribution of scores is almost exactly the same in these 'true' onset – nucleus pairs as in the comparison between first and last lexical item. Overall, the nuclei were deemed to be slightly more prominent than the onsets. The difference is significant ($p < 0.05$) for both groups of raters, but still very small: 0.12 degree of prominence for the Danish raters and 0.07 for the English raters. In individual utterances the difference is significant in six utterances for the Danish raters but only in one utterance for the English raters. This difference between the rater groups is mainly caused by the difference in group size and thereby the number of observations; ten the Danish raters and only six the English raters, which means that only large differences will be significant in the latter group. The general distribution of ratings for onsets and nuclei becomes clearer in the diagrams in Figure 6.4 below.

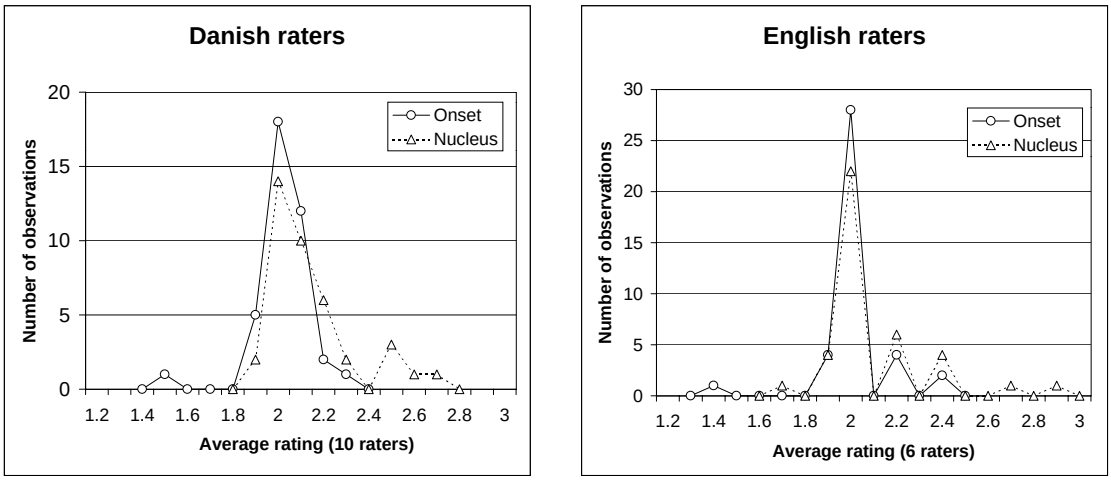


Figure 6.4. Distribution of ratings for onsets and nuclei. Onset ratings are indicated by solid lines and circles, and nucleus ratings are indicated by dashed lines and triangles. Interval width for the histograms was 0.1.

The general pattern of distribution is more or less the same for the two groups of raters: most of the scores are centred around the value 2, indicating normal, full stress. This is true for both onsets and nuclei, but they differ with regard to deviation from the central tendency. Only one stressed word, in onset position, was deemed considerably less prominent than 2, and more nuclei than onsets were deemed considerably more prominent than 2. This confirms the conclusion from Test 1 that the statistical significance between the prominence ratings of onsets and nuclei (or first and last lexical item) is caused by a few individual items where this pattern is very clear, rather than by a general tendency for nuclear syllables to be more prominent than onset syllables. Making the nucleus very prominent seems to be a strategy which is available to speakers but which does not appear to be obligatory, contrary to the common descriptions of the nucleus as the most prominent syllable in the intonation unit. Of course, this conclusion is only valid for neutral, context-free

utterances where no word has been emphasised for focal contrast or other semantic or pragmatic reasons.

6.8.1.2 Intervening lexical items

Intervening stressed words between the first and last lexical item have scores which are fairly similar to the scores in the other two tests, although there are some notable exceptions. The item immediately following the onset has been deemed considerably less prominent than the surrounding items in most cases, but the general tendency for a strong – weak alternation pattern is much less clear than in Test 1 and Test 2. That is closely connected with the fact that the perception of a nucleus (plus a following boundary) in this experiment is very different from the perception of strong stress in the other tests, even though both result in the score 3. This is reflected clearly in the ratings of the sentence *dsi*. In speaker 2F's version of the sentence all four raters indicated a primary accent/nucleus on the word 'insist' giving an average rating of 3. In comparison the raters in Tests 1 and 2 gave average ratings of 1.9 and 2.0 respectively, deeming this word to be exactly as prominent as the first lexical item in the utterance, 'Stalin'. In the other three representations of this utterance in the test there is good agreement among the raters across all three tests, but it seems obvious that fairly weak boundary (or other) cues can and will have a much larger effect on prominence ratings within a system such as the British school of intonation than within a simpler system such as the one used in Test 1 and Test 2.

The overall ratings provide no clear answer to the question about whether intervening items have secondary or tertiary stress – a choice which is expected to depend on the size of the F_0 downstep. The ratings are generally between 1.5 and 2, and this is also the case for individual utterances (not listed here). The disagreement among the raters is quite large about these items, and it thus seems that the two fundamentally different types of contour are difficult to identify and distinguish consistently.

6.8.2 Marked information structure – British tradition

Below are the ratings for each sentence with a marked information structure, across all speakers.

Focus	*Sentence*
f1	$^{3.00}$Paul $^{1.00}$sings
f2	$^{1.54}$Paul $^{3.00}$sings
f1	$^{2.96}$Bill $^{0.46}$struck $^{1.33}$Ann
f2	$^{1.62}$Bill $^{3.00}$struck $^{1.17}$Ann
f3	$^{2.20}$Bill $^{1.05}$struck $^{3.00}$Ann
f1	$^{2.95}$Jane $^{0.75}$kissed $^{1.00}$Frank $^{1.00}$tenderly
f2	$^{1.69}$Jane $^{3.00}$kissed $^{0.75}$Frank $^{1.00}$tenderly

Focus	*Sentence*
f3	$^{2.00}$Jane $^{0.75}$kissed $^{3.00}$Frank $^{1.25}$tenderly
f4	$^{1.90}$Jane $^{1.00}$kissed $^{1.75}$Frank $^{3.00}$tenderly
f1	$^{0.00}$The $^{3.00}$party $^{0.00}$was $^{1.00}$cancelled
f2	$^{0.00}$The $^{2.08}$party $^{0.00}$was $^{3.00}$cancelled
f1	$^{0.00}$The $^{3.00}$cook $^{0.00}$was $^{1.04}$smelling $^{0.00}$the $^{1.08}$soup
f2	$^{0.00}$The $^{1.92}$cook $^{0.00}$was $^{3.00}$smelling $^{0.00}$the $^{1.00}$soup
f3	$^{0.00}$The $^{2.12}$cook $^{0.00}$was $^{1.67}$smelling $^{0.00}$the $^{3.00}$soup
f1	$^{2.95}$Sheila $^{1.05}$examined $^{0.00}$the $^{1.00}$patient $^{1.10}$carefully
f2	$^{1.75}$Sheila $^{3.00}$examined $^{0.06}$the $^{1.00}$patient $^{0.94}$carefully
f3	$^{2.00}$Sheila $^{1.19}$examined $^{0.00}$the $^{3.00}$patient $^{1.00}$carefully
f4	$^{2.17}$Sheila $^{1.46}$examined $^{0.00}$the $^{1.46}$patient $^{3.00}$carefully
f1	$^{0.00}$Is $^{2.95}$Peter $^{0.00}$a $^{1.05}$doctor $^{0.00}$in $^{1.50}$Paris
f1	$^{0.30}$Did $^{2.90}$Stalin $^{1.05}$insist $^{0.00}$on $^{0.00}$an $^{1.15}$equal $^{1.10}$distribution $^{0.00}$of $^{2.05}$wealth

The ratings of the lexical words are plotted in Figure 6.5.

6.8.2.1 Problems concerning individual words

An analysis of disagreement about individual words reveals that the mean scores of some items deserve a comment. There was general disagreement about the word 'Bill' in *bsa (f2)* in the utterances by all six speakers. The disagreement is distributed over all four categories, but is centred around 1 and 2. The overall mean value is therefore perhaps not inappropriate as an indication of the central tendency.

In sentences *ps (f2) bsa (f3)*, and *jkft (f4)* there is some disagreement about whether the words 'Paul, struck' and 'kissed', respectively, in pre-focal position have secondary or tertiary stress. In addition, in each of those sentences one rater deemed these words to be completely unstressed, which influences the overall ratings. All three sentences exhibit stress clash, and especially the words in second position, *struck* and *kissed* have been shown before (Test 2) to be very sensitive to reductions, both in neutral utterances and in utterances with an explicit focus. The present results again indicate categorically different perception – either full, normal stress or no stress – among the raters in these cases. It is difficult to say whether the arithmetic mean value is a good expression of the central tendency in the light of this disagreement, especially with so few raters. Since most of the disagreement about these words concerns the two (adjacent) categories secondary and tertiary stress they have not been excluded from analysis, but it is worth noting that the average rating of 1 covers some indeterminacy between 0 and 2.

Finally, in *pdp* and *dsi*, with a focus on the first lexical item, most raters indicated either tertiary stress or primary stress on the last lexical item, 'Paris' and 'wealth', respectively, and only rarely secondary stress, as the average ratings of (around) 2

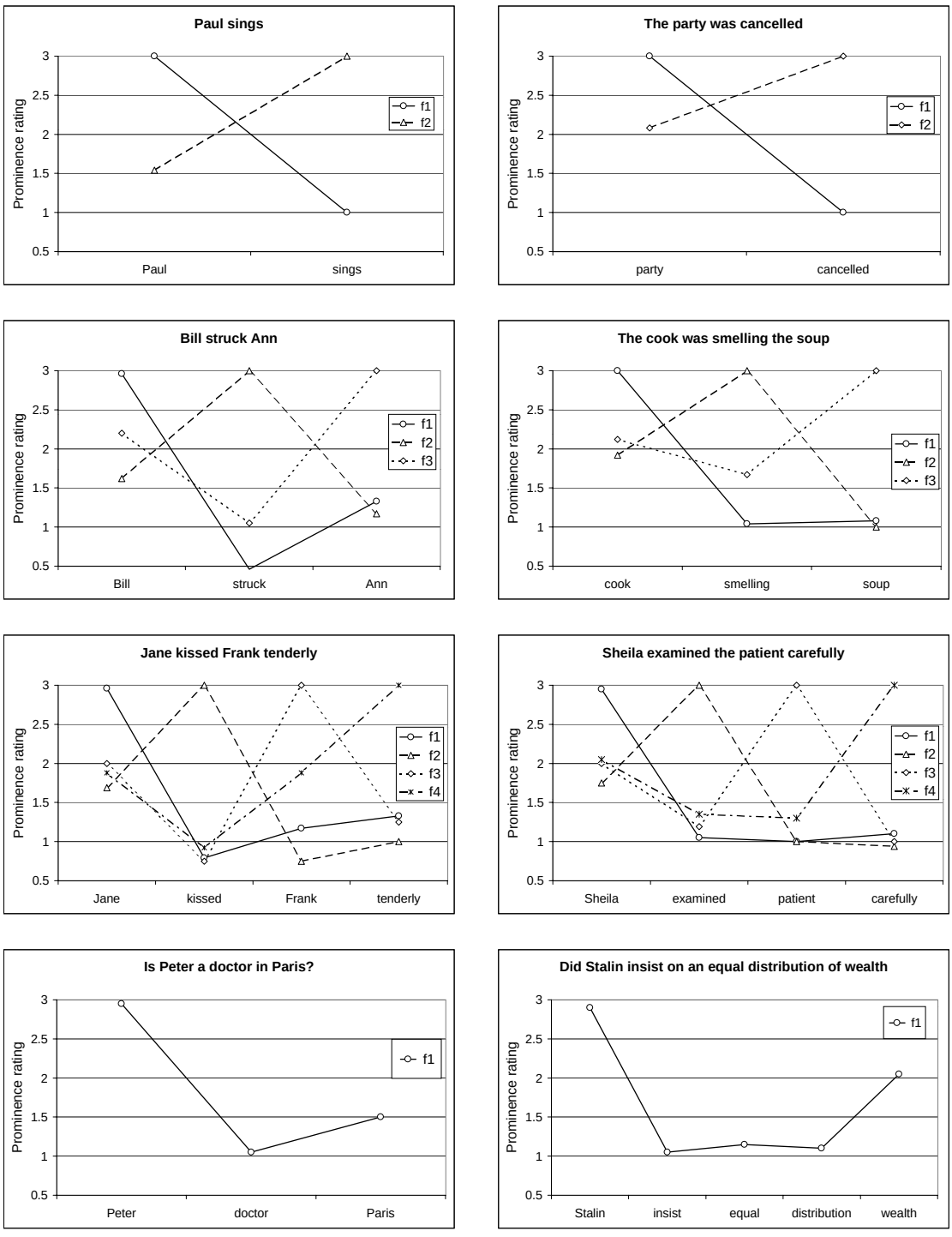


Figure 6.5. British school of intonation. Prominence ratings of the stressed syllables in sentences with a marked information structure. The labels *f1, f2, f3, f4* indicate that the position of the intended focus is on the first, second, third and fourth stressed syllable, respectively.

might suggest. In fact, the choice of 2 (secondary stress/accent) might strictly speaking not even be allowed according to the British framework, which normally states

that post-nuclear syllables are unstressed or have tertiary stress, that is, no tonal prominence. But as mentioned, there is one particularly problematic case, namely the fall-rise in early position in the intonation unit. The 'rise'-part of this complex nuclear tone falls at the end of the intonation unit, and thereby lends some tonal prominence to the syllable with the rise. Seven or eight (one case was difficult to determine) of the ten utterances had this tonal configuration, and were the ones which caused disagreement. The two utterances which had either a fall or a rise (one of each) yielded full agreement among the raters. The interpretation of the final rise is normally regarded as a choice between nucleus (primary accent) and part of the tail (tertiary stress), see Section 6.4.

Not all representatives of the British school of intonation analysis draw the same conclusions about this problem, and Cruttenden (1997) states that 'It seems to be generally true in English that a final accent dependent on a rise following a fall is normally downgraded from its status as nucleus' (p. 43). However, it is not entirely clear how far down in the stress hierarchy it has to be downgraded. One rater – *r* 4 – used tonetic stress marks in the test and marked some rises as 'low rises'. Since the ratings had to be coded numerically the rater was asked to indicate his perception of the stress level on those words, and his response, referring to the sentence 'Did Stalin insist on an equal distribution of wealth?', was:

> It would have to be 2 [i.e. secondary stress]. ″-sist″, ″equ-″and ″-bu-″ are more prominent than 0, and ″wealth″ is less prominent than ″Sta-″, so.....

In other words, 'wealth' was felt to be more prominent than the other post-nuclear stressed words, but less prominent than the nucleus. If the system had allowed for the difference between a default nucleus and emphatic stress on a word with narrow focus, 'wealth' might be considered a nucleus and still fit into the description given by the rater. As it is, some concept of phrasal subordination is needed to account for the information structure in the utterance if both 'Stalin' and 'wealth' are labelled as nuclei, but with a more important, or superordinate, informational focus on *Stalin*. The scores from the other raters were almost evenly distributed between 1 and 3; so although the mean values do not really indicate the typical response, they do perhaps express the average perception fairly well. At least, the scores are quite similar to the ones in Tests 1 and 2.

6.8.2.2 Comparison with Tests 1 and 2

Comparing the ratings with the results from Tests 1 and 2 reveals more similarities than differences. To start with the latter, the ratings for post-nuclear (or post-focal) items are around 1, which is somewhat lower than in Test 2 (English raters) and especially Test 1 (Danish raters). This is obviously related to the fact (as mentioned above) that the framework does not normally allow levels of prominence higher than tertiary stress in the tail of an intonation unit, so an indication of a higher level of stress is linked to the perception of another nucleus (see above for an exception). The raters indicated this interpretation in only a few cases.

Apart from this the ratings are surprisingly similar to the results in Tests 1 and 2, both in terms of the variation according to different positions of the pragmatic focus, and in terms of actual numerical values on the four-point scale. I have already mentioned that the stress level on post-focal items is generally around 1, except when the final item in an utterance is perceived as a (second) nucleus. The stress level on pre-focal items is reduced in inverse proportion to their distance from the focused item, just as in the other two tests. Only a few of the differences related to this observation (for example, 'Bill' in *bsa (f2)* versus 'Bill' in *bsa (f3)*) are statistically significant, partly because of the low number of observations – 16–24 per context – and partly because the variance is higher than in Tests 1 and 2 as a result of the (occasional) indeterminacy between the categories 'tertiary stress' and 'nuclear stress'. But the variations in perceived prominence (or stress) are the same in all three tests, which is a strong indication that the observed differences are not random.

6.8.2.3 Default nucleus versus final focus

The most apparent difference between the results from the traditional British style ratings and the prominence ratings from the previous experiments is in the distinction between neutral, context-free utterances and utterances with an explicit focus on the final lexical item. In Tests 1 and 2 this distinction was marked very clearly in the perceived prominence on the final item, which was around 2 (normal stress) in neutral utterances and almost 3 (strong stress) in utterances with focus on this item, and also in a small but significant reduction in prominence on the pre-focal items. In this experiment the final lexical items are identified as primary accents (nuclei) in both these contexts, and therefore have the same rating, namely 3. The notational system of the British school does not allow for a distinction between a default nucleus, and a nucleus with emphasis for contrast or other focus, at least not in the notation of the nucleus itself. Any differences between the two contexts will therefore have to be represented in the notation of the pre-focal words, if it is present at all. There are six sentences where the two contexts can be compared: *ps, pc, bsa, css, jkft* and *sepc*. Visual inspection of the diagrams in Figure 6.2 and Figure 6.5 shows that sentences *bsa (n) (neutral)* and *bsa (f3) (focus on third item)* are identical in terms of prominence. In the five other sentence pairs there is a difference in prominence on at least some of the pre-focal items. In sentences with two lexical items the one pre-focal item is less prominent in the version with an explicit focus than in the neutral version. The difference is significant for sentence *ps* ($p < 0.05$) but not for *pc*. In sentence *css*, with three lexical items, both pre-focal items have lower prominence in the (*f3*) version, but the difference is only significant for the first item. Finally, in sentences *jkft* and *sepc*, with four lexical items, the prominence of the second item is significantly lower (0.4–0.65 on the four-point scale, $p < 0.05$) in the (*f4*) version, whereas the differences are smaller (0–0.25) and non-significant for the first and third lexical items. As mentioned before, it is difficult to test observed differences in this material because of the small number of raters, but the regular pattern of variation found here at least indicates a strong tendency for reduction on pre-focal items,

similar to what was observed in Tests 1 and 2. In other words, in the traditional British system differences between neutral utterances with broad focus and utterances with narrow focus on the final word are indicated in the ratings of pre-focal items only.

6.8.3 The British system and prominence ratings

The purpose of this test was to examine the notational system of the British framework, and to see if the stress and accent hierarchy of this framework is an appropriate basis for an investigation of prominence relations in British English. Descriptions of the British system often make reference to prominence in the definition of the various stress/accent levels, and specify which phonetic features are associated with them (Cruttenden 1997: 44, Gimson 1989: 270). It could therefore be assumed that the stress/accent hierarchy would correlate well with a simpler, less theoretically determined, notion of prominence as used in Test 1 and 2. In order to compare the results across both systems, or scales, it was assumed, for the sake of argument, that the four scale points of each system could be equated, although it was noted that such an equation was not completely justified by the description of the British framework which was used. For one thing, equating the primary accent of the British system with (extra) strong stress might be inappropriate, since the label (extra) strong was expected to imply some sort of emphasis, which is not an inherent part of a primary accent.

This suspicion was confirmed by the results. The final lexical items of neutral utterances were consistently identified as primary accents, or nuclei, and achieved the rating 3. In contrast, such items only achieved an average rating of around 2.1 in the other two tests. This does not in itself present a problem, since it might just indicate a different use of the scale, but two further observations complicate matters. First, in utterances with an explicit focus on the final lexical item the raters using the British framework again identified this item as a primary accent (3), indicating no difference between such items and the final items of neutral utterances. In Tests 1 and 2 on the other hand, final focused items received prominence ratings close to 3, indicating a large difference between the focused items and the same words in a neutral utterance. So the difference which could be predicted from the shift in information structure was treated very differently in the two systems. Secondly, the first and last lexical items were deemed almost equally prominent in Tests 1 and 2, at just over 2 on the scale. In Test 3, using the British system, the first items received ratings of just over 2.2, which although slightly higher than in the other tests is much lower than the final items. This reflects the fact that most of the first lexical items were heard as onsets, which by definition have secondary accent/stress (2). The observed difference between first and last lexical item is therefore just as predicted, but was not found in the simpler prominence ratings in Tests 1 and 2. These differences between the two systems have implications for the expectations one may have about the acoustic manifestation of the (prominence on the) first and last items. The results from Tests 1 and 2 predict no significant differences in features which are

linked directly to perceived prominence, while the results from Test 3 do predict such a difference.

The disagreements about stress level on post-nuclear items is another complicating factor. Because of restrictions imposed by the model a rater must choose between either no stress, tertiary stress or a new primary accent in this position, and in the case of an early/advanced nucleus with a falling-rising tune there was very large disagreement about the latter two categories. Cruttenden (1997) devotes much attention to the problem of identifying such falling-rising tunes, or rather distinguishing between a single falling-rising tune and a combination of a falling tune plus a rising tune in a new intonation unit. This problem was also commented on by one of the raters in this experiment and, as stated, confirmed in the high level of disagreement about these utterances. It is of course a theoretical consideration whether one sees this as a choice between two non-adjacent categories in the hierarchy, or whether a 'middle-ground' solution can be allowed, as the one taken by one rater who preferred to assign secondary stress to the post-nuclear rise, but it is an undeniable problem if one wants a reliable and consistent indication of stress (and thereby prominence) levels in English utterances. If a large number of raters are used, the different interpretations of the tonal configurations may even each other out, and the mean value of the scores might be considered a good expression of the central tendency. But with a small number of raters one might easily end up with large differences in mean values between very similar items. These differences would then perhaps be very difficult to justify on the basis of subsequent acoustic analyses of the material.

All in all it must be concluded that the stress/accent hierarchy of the British school of intonation, as defined by Cruttenden (1997) and Gimson (1989), is a problematic frame of reference for investigations about prominence levels in English. Some of the predictions about stress levels made by this model and also confirmed in Test 3 here differ markedly from the prominence ratings obtained using a more theoretically agnostic system such as the one in Tests 1 and 2. Of course, that might testify to the inadequacy of the system used in those tests, but I believe that I have pointed to specific weaknesses in the traditional British framework above.

Some might find my criticism of this framework unjustified since I have in fact taken that system, bent it to suit my own purpose (for which it is not primarily designed) and then dismissed it because it does not yield the same results as some other framework. But it has not been my purpose here to dismiss the British system as a model of intonation, or a general descriptive framework of British English intonation. I have only been concerned with the stress and accent hierarchy of the model, and the references to prominence in the definition of this hierarchy. The establishment of the hierarchy may well be based on observations of utterances for which it seems very appropriate, or on valid intuitions about the hierarchical nature of stress levels, but the restrictions the model imposes on the possible stress level of a syllable because of the syntagmatic rules for the structure of an intonation unit is problematic for the analysis of some types of sentence, for example some neutral, context-free

utterances. One of these problems might be explained as follows: it has often been observed, or stated, that in most utterances one syllable stands out as more prominent than the others and is characterised by strong pitch excursions, longer duration and stronger intensity than other syllables in that utterance (in addition to full vowel quality). This syllable has been designated the nucleus of the utterance (or more correctly, the intonation unit), and the generalisation has been made that all utterances/intonation units must contain one (and only one) nucleus. However, in the absence of any strong phonetic cues about the location of the nucleus the last lexical item of the unit will be perceived as carrying nuclear stress (Brown *et al.* 1980: 145-146). In such cases the identification of the nucleus (as the last lexical item) may depend as much on expectations about the information structure of the utterance as on the prominence relations within it. But if a final nucleus is not necessarily more prominent than the other syllables in the utterance, and extra prominence on a final nucleus does not necessarily signal (narrow) focus or emphasis on this item, then these two cases may be functionally equivalent. However, since extra prominence on a final item may in fact also be associated with narrow focus, the question becomes what (else) is required before the listener will choose this interpretation over broad focus. As a first step in this direction it would be helpful to have experimental verification of the perceived information structure in the utterance. Do the listeners interpret the utterances in the way that (it is assumed) the speakers intended, and if there are mismatches between intended and perceived structure then where do these occur? In Chapter 7 I present a small experiment which examines these questions.

6.9 High preheads – accented or not?

In the presentation of the prominence ratings in Tests 1–3 some words stood out from the general pattern that grammatical words are completely unstressed, namely the words 'Is' and 'Did' from sentences *pdp* and *dsi*. While most of the 18 utterances of *pdp* and *dsi* which were included in the perception experiments displayed this pattern to some extent, there were five utterances in which it was particularly clear, as shown in Table 6.6.

			Prominence ratings		
Sent	*Spk*	*Word*	*Test 1*	*Test 2*	*Test 3*
pdp n	4M	*Is*	1.5	0.67	1.5
dsi n	4M	*Did*	2.2	1.0	1.5
dsi n	6M	*Did*	0.8	1.33	1.25
dsi n	2F	*Did*	2.5	1.33	1.5
dsi f1	2F	*Did*	1.5	0.5	1.0

Table 6.6. Five utterances in which the initial, grammatical word is deemed prominent by the raters. Prominence values are from Tests 1–3.

Four of the five utterances are neutral, context-free versions, but one – *dsi f1* 2F – has a (narrow) focus on the immediately following word, 'Stalin'. It appears that the Danish raters in Test 1 generally perceived 'Is' and 'Did' in these utterances as more prominent than did the native English raters in Test 2 (except in utterance *dsi n* 6M). Examining the ratings of the individual listeners and utterances reveals a larger dispersion in Test 1 than in Test 2. The Danish raters varied between hearing categories 0, 1 and 2 in some utterances and 2, and 3, or even 0, 2 and 3 in others, while the English raters all heard either category 0 or 2, that is, no stress, or normal, full stress. This apparent difference between the Danish and English raters may be connected with the tonal properties of the utterances. This aspect will be treated below, but see Figure 6.6 (1) and (2) for exemplifications. The ratings from Test 3 (British school) are not directly comparable with those from the other two tests, but confirms the fact that these words were not consistently heard as completely unstressed. In Test 3 they were generally heard as either unstressed or as carrying secondary accent and in one case even a primary accent.

The above-mentioned large dispersion in the ratings means that these grammatical words were among those which caused the most disagreement among the raters. Not only in the pairwise comparisons, but also, and notably, in the magnitude of the disagreements as expressed by the standard deviations. These are presented in Table 6.7 together with the mean standard deviations for the whole material in the three tests. Note that value of (roughly) 1.10, 1.09 and 1.06 standard deviation units characterise random variation in Tests 1, 2 and 3, respectively.

			Disagreements		
Sent	*Spk*	*Word*	*Test 1*	*Test 2*	*Test 3*
pdp n	4M	*Is*	1.17	1.07	1.00
dsi n	4M	*Did*	0.84	1.20	1.00
dsi n	6M	*Did*	0.62	1.07	2.25
dsi n	2F	*Did*	0.28	1.07	1.00
dsi f1	2F	*Did*	0.50	0.70	1.33
Mean S.d. (all words)			0.29	0.33	0.18

Table 6.7. Inter-rater disagreement about the two prominent grammatical words, expressed as the standard deviation of the prominence scores. The grand mean standard deviation of all words is included for reference.

The standard deviation values are clearly very high; many of them are close to or even higher than the random variation values. This is a good indication that the words did not just cause some uncertainty among the raters about two adjacent categories, but rather that they are perceived in categorically different ways. One might say that these words are indeterminate, or perhaps indeterminable, with regard to accentuation (or stress level): some perceive them as stressed – and in that case often as fully (or even strongly) stressed rather than partially stressed – while others hear them as

completely unstressed. At this point I can only offer a speculative attempt at an explanation (backed by a few observations): it is possible that these words are very prominent in a strict acoustic and perceptual manner, but that they are unlikely to be identified as stressed or accented in the functional sense of 'marking important words' or signalling 'semantic peaks' because of (1) their status as grammatical words and/or (2) certain conflicting acoustic cues – for example tonal versus rhythmic prominence – or their integration into the intonational pattern.

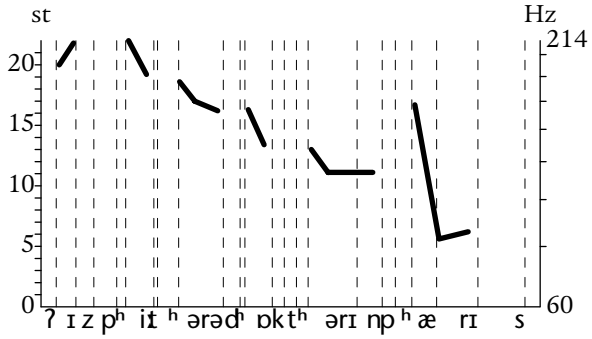
With regard to the latter speculative statement it is worth noting that all five words are produced on a high pitch, that is, with F_0 values at the top end of the speaker's range. Figure 6.6 shows simplified traces of F_0 (and duration) for two of the utterances and, for comparison, for utterance *pdp n* 6M in which 'Is' was not perceived as prominent. Note that the traces in Figure 6.6 are not averaged traces, as in Section 2.5.2, but simplified versions of individual utterances.

As it appears from traces (1) and (2) in Figure 6.6, as well as the traces for the other three utterances in Appendix B, Section B.6, the words in question are placed very high in the F_0 range – as high as, or even higher than, the following lexical word. In contrast, in utterance (3) the initial word 'Is', which was not deemed very prominent, has much lower F_0 – lower in the speaker's range and lower (by 4 semitones) than the following word. The high pitch on these initial grammatical words may be (part of) the explanation of why the Danish raters generally perceived them as more prominent than did the English raters. In Danish the first jump up from the baseline is normally associated with the first stressed syllable, and (initial) unstressed syllables high in the F_0 range are quite rare, or confined to affected speech styles. Danish raters may therefore have a strong assumption that such high-pitched syllables are stressed – a stronger assumption, so it seems, than do English raters.

In (3) the grammatical word holds the position in the tone group which is referred to as the *(low) prehead* in O'Connor and Arnold (1973: 22) – the low-pitched unaccented syllable(s) which precede(s) the *head*, which begins with the first accented syllable. (An *accent* in their system is a stressed syllable which is made prominent *inter alia* by pitch). The O'Connor and Arnold system also allows for *high preheads*, that is, syllables which carry no stress or tertiary stress and which precede the head. It is thus possible to regard the fairly prominent grammatical words in the above utterances as high preheads, which makes them functionally equivalent to the low prehead in *pdp n* 6M. It is somewhat problematic to determine whether these prominent words with high F_0 should be regarded as high stressed preheads or as the first accent (the 'onset' in the terminology adopted here from Knowles 1987) based on O'Connor and Arnold definitions (a schism which is reflected by the categorically different prominence ratings). The differences lie in the combinations of types of prehead and types of head, but in certain configurations the two possible solutions may look identical in terms of pitch variation.

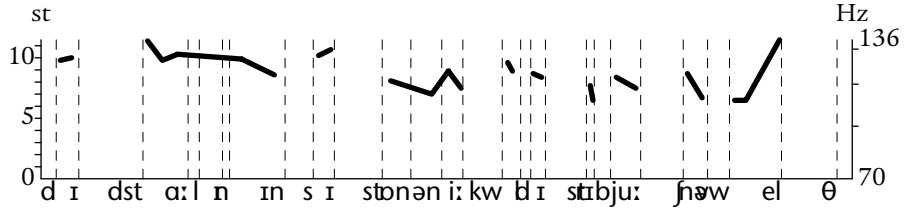
Stronger evidence of the status of these words can be found in the prominence ratings in Tests 1–3. It was noted in the quote from Knowles (1987) and confirmed in Tests 1–3 that stressed syllables between the onset and the nucleus tend to be

(1)



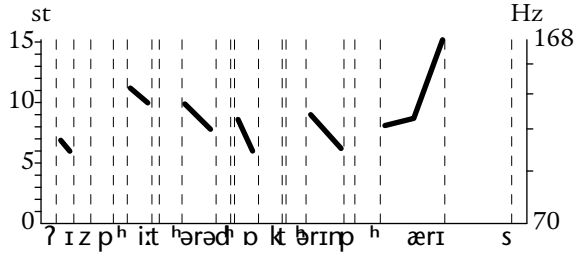
Sent *pdp n* Spk 4M

(2)



Sent *dsi n* Spk 6M

(3)



Sent *pdp n* Spk 6M

Figure 6.6. Simplified F_0 tracings of two utterances with a prominent initial, grammatical word. Utterance (3) – *pdp n* 6M – with low perceived prominence on 'Is' is shown for comparison. See the other three utterances in Appendix B, Section B.6.

deaccented to some extent. In particular, the lexical item immediately following the onset was often deemed considerably less prominent than the onset. But the initial, grammatical words 'Is' and 'Did' in the above five utterances were all deemed less prominent than the following word in Tests 2 and 3 and in three out of five cases in Test 1 (Danish raters). This speaks clearly against an interpretation of these words as onsets. The strong – weak alternating pattern which was observed in most of the sentences only works in these utterances if it starts with the first lexical item – regardless of the prominence level of the grammatical word that precedes it.

CHAPTER 7

Test 4 – Perceived information structure

7.1 Introduction

Tests 1 and 2 demonstrated a clear difference between utterances with a single highlighted constituent, in which one element was consistently deemed to be more prominent than the others, and neutral, context-free utterances, in which there was often not one item which stood out as the most prominent one. In 15–20% of the neutral utterances the final lexical item was more prominent than the others, but in the remaining 80–85% there was no significant perceived difference between the first and last lexical items. In a few cases the first item was deemed more prominent than the last item. This raises some questions about the perception of information structure in the utterances in my investigation:

(1) How strong is the connection between variations in prominence relations and perceived information structure? Put differently, to what extent can the perceived information structure be predicted from the prominence ratings obtained in Tests 1 and 2?

(2) Can listeners distinguish clearly and consistently between (the relatively few occurrences of) neutral utterances in which the final item is deemed more prominent than the others, and utterances where the final element has been highlighted?

(3) It must be assumed that listeners will perceive an informational focus on items which were highlighted by the speakers as a response to a question about a specific item in the utterance. However, in some neutral utterances a non-final item was deemed to be the most prominent one. Are such items also likely to be perceived as highlighted or is this prevented by other properties of the utterance?

To examine these questions a small experiment was set up, in which selected utterances from the preceding tests were presented to a group of listeners who were asked, indirectly, to indicate their interpretation of the information structure of the utterances.

7.2 Method

Thirty-five of the 183 utterances which were used in Tests 1–3 were selected, most of them from Group 1 (see Section 2.2.1), that is, sentences which were found both in a neutral, context-free version and in versions where a specific item had been

emphasised as a response to a question about an item in the sentence. This included a few of the utterances which were excluded from analysis in the previous experiment because they deviated from the norm, in particular, some neutral utterances by speaker 3F which were characterised by having a high degree of perceived prominence on the first lexical item.

The idea was to present each utterance to the listeners and ask them to identify the context in which the utterance had been produced. The possible contexts were those which had originally been used to elicit utterances with an explicit focus (see Section 2.2.1) or a context-free reading. For example, a version of the sentence *bsa* was played to the listeners who would then indicate which one of the following contexts applied:

Bill struck Ann.
[ ] No question.
[ ] Who did?
[ ] He did what to her?
[ ] Who did he strike?

The responses to these questions say nothing about perceived prominence levels, but only about the interpretation of the utterances in terms of information structure (chosen from a restricted set). It is really a test of whether the intended information structure is the same as the perceived information structure. If all the listeners are able to decode all of the contexts 'correctly', that is, if there is full agreement between intended and perceived message, the results will be of restricted value, except as a confirmation that the speakers were able to get their message across, and that the prominence ratings obtained in the previous tests did not represent prosodically deviant utterances. But if there are systematic mismatches between intended and perceived structure, or inter-listener disagreement about certain utterances, then that may reveal information about prosodic prerequisites for marking an utterance as neutral or as having an explicit focus. This knowledge might also be useful in any subsequent attempt to establish the acoustic correlates of prominence and perceived information structure and the relation between the two.

7.3 Test setup

7.3.1 Internet test

The Internet test consisted of a small computer program (a CGI-script), which presented the utterances one by one together with checkboxes for the appropriate contexts. The subjects could listen to an utterance as many times as they wanted before choosing a context. They submitted their answers by pressing a 'send' button, which would also present the next utterance.

There was an online tutorial which explained how the utterances had been collected, including samples from the original recording sessions and a few practice utterances, to enable the listeners to get acquainted with the process. The test and instructions can be found online at http://www.cphling.dk/pers/chrjen/is/.

7.3.2 Recorded test

It proved impossible to recruit as many listeners for the online test as required, and therefore a more traditional test was prepared, with a fixed number of repetitions and fixed intervals between repetitions. The advantage of such a test is that it can be tape-recorded and played in a classroom, a language lab or in some other suitable location. The order of the test items was the same as in the online version; each utterance was played back three times with two second intervals. There was a five second pause before each new utterance. The listeners indicated their responses on answer sheets with the same options for each utterance as in the online version. A small 'beep' was used to indicate the end of a page. The instructions to this test were based on the online version, and the explanation of the procedure was recorded together with the examples for demonstration to ensure that they would be constant across different test sessions. The listeners were also given a printed version of the instructions. Before the test began they were asked if they had understood the instructions and knew what they were expected to do. One test session ran over a loudspeaker in a classroom, and three other sessions took place in a language lab using high quality headphones.

7.4 Subjects

The online test was performed by four members of staff and one post-graduate student from the English Department at the University of Copenhagen, none of whom is a linguist/phonetician. The fixed format test was performed by 57 listeners; all except two were first-year students of English at the University of Copenhagen, at which point they will have received (at least) 7–9 years of English language training. One respondent was a phonetics teacher at the English Department and one a technician with good English skills. Most of the listeners were L1 speakers of Danish, but there were also speakers with Faroese (3), English (2), Icelandic (1), Romanian (1), Turkish (1), or mixed German/Danish (1) or Japanese/Danish (1) language backgrounds. The significance of the different listener backgrounds and test setups is treated below.

The listeners did not report any problems understanding the task, and since the duration of the test was only about ten minutes they showed no signs of fatigue during or after the test. The three repetitions of each item in the fixed format test seemed sufficient to let the listeners decide on a context without straining them.

7.5 Results

7.5.1 Listener reliability and agreement

Of the 62 listeners who participated in the experiment four were excluded because they did not choose a context for every item, or because they gave two answers to the same utterance. The reliability of the remaining responses and agreement between listeners was tested using two fairly simple χ^2 'goodness-of-fit' tests: one for inter-listener agreement about each utterance, and one for each single listener's ability to

identify the intended contexts. The two tests are similar and rely on the calculation of the chance probability of obtaining the correct answer, that is, a match between intended and perceived information structure, for each utterance. The number of categories (possible contexts) varies between three[1] and five (number of lexical items in the utterance + one for neutral, or no context), and the expected chance proportion of correct answers is therefore 1 divided by the number of categories. For example, for the utterance 'Paul sings' there are three categories, so the expected chance proportion of correct answers is 1/3, or 0.33. When applying the χ^2 test to each utterance the expected number of correct answers for the whole group of listeners will be 0.33 × number of listeners, which can then be tested against the category which received the highest number of scores. A small note about agreement and correctness is needed here. If all listeners agree that the utterance *pc (n)*, that is, a neutral version of 'The party was cancelled' was uttered in response to the question 'The match was?', which implies focus on the first lexical item, then agreement is high, but correctness for the whole group is low. When testing whether the answers are random or not the relevant measure is agreement, and therefore the category which received most scores is used (here *f1* = 100%), but in the analysis of the scores below, the appropriate measure will most often be correctness, or whether the listeners were able to decode the intended information structure of the utterances. In such cases the number of expected correct answers is tested against the number of observed correct answers. The expected proportion of incorrect answers is simply the residual of the above calculation, or 1 – (1/number of listeners).

When testing the performance of a single listener the total proportions of expected correct and incorrect answers can be tested against the total number of correct and incorrect answers for that listener. According to this procedure two of the listeners failed to perform at better than chance level ($p > 0.05$), and their scores were consequently excluded, which leaves 56 sets of responses for further analysis.

In all the utterances inter-listener agreement is significant ($p < 0.05$), but in three cases listeners agreed on a different context than the one intended (or so it must be assumed) by the speaker. Testing inter-listener agreement about individual utterances has a different purpose than testing intra-listener correctness, since the failure to achieve agreement about an utterance or failure to identify the intended context does not mean that this utterance should be excluded. In fact, such utterances are at the centre of interest in this experiment, provided of course that there is some level of overall agreement from which they form a deviation. The results of the χ^2 test show that this is indeed the case, and it must be concluded that the task could be solved with at least some confidence by all except two listeners, and that there was agreement beyond chance level about all utterances, although not always in favour of the 'correct' answer.

[1] There were only two possible contexts for sentence *pdp*, namely neutral and focus on the first item, see Section 2.2.1.

7.5.2 Overall identification of contexts

Not all listeners were equally successful in identifying the intended information structure of the utterances. The distribution of the number of correctly identified contexts for the 56 listeners is shown in Figure 7.1.

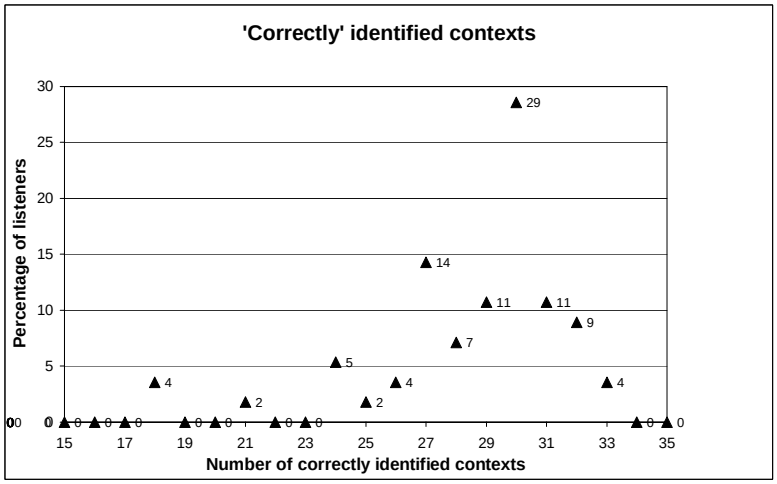


Figure 7.1. Distribution of the number of correctly identified contexts for 56 listeners. The numbers on the x-axis refer to the number of correctly identified utterances/contexts, not to specific utterances. It thus appears from the diagram that 14% of the listeners identified exactly 27 utterances/contexts of 35 possible and 29% identified exactly 30 contexts.

Just over half of the listeners (52%[2]) were able to identify 30 or more of the 35 contexts, and almost 90% identified 25 or more. There is a clear peak around 30, which is by far the most frequent number of identified contexts, with half of the listeners recognising 30 contexts plus/minus one. The expected random number of correct identifications (for example, by guessing) is 8.9, so most of the listeners are clearly far above this threshold (they are all above chance level, see the previous section).

None of the listeners were able to identify all of the contexts correctly; the highest score was 33, which was accomplished by two listeners (4%). So although most of the listeners could perform the task with a good deal of confidence, they could all be misled on at least a couple of occasions. The analysis below explores whether this was caused by a general level of uncertainty about most items, or by a few deviant utterances which were difficult for most listeners.

7.5.3 Differences between listener groups

The 56 listeners fall into different groups depending on factors such as the test setup and linguistic background (see above), and it is possible that there are systematic differences in the ratings of these groups, either in terms of the average number of 'errors' that the members of each group make, or in the distribution of these errors.

[2] Adding up the relevant numbers in Figure 7.1 yields 53% due to rounding errors.

Furthermore, it is possible that the listeners behave in systematically different ways irrespective of their linguistic background or the test setup, that is, that some listeners respond to certain phonetic cues while others respond to other cues. This would give rise to groupings within the whole group of listeners which could only be determined by analysing response patterns. Such systematic differences may obscure otherwise clear tendencies when all listener responses are grouped, and consequently certain tests were performed in an attempt to find patterns in the scores of the various sub-groups that may prohibit such an overall grouping.

7.5.3.1 Average number of errors

There were three different sets of test conditions, or setups, which may have had an influence on the results. The groups are of very unequal sizes, with five listeners performing the test from the Internet, 16 in a classroom via a loudspeaker, and 35 in a language laboratory using headphones. As shown in Figure 7.2 and Table 7.1, the number of correctly identified contexts varied somewhat between the groups:

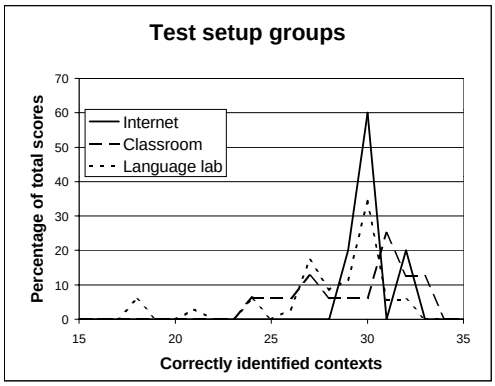


Figure 7.2. Distribution of the number of correct contexts in the three test setups.

Test setup	*N*	*Correct*
Internet	5	30.2
Classroom	16	29.4
Language lab	35	28.0

N = number of listeners in the group

Table 7.1. Number of identified contexts in each test setup (grand mean).

The difference between the Internet group and the language lab group is significant ($p < 0.01$, two-tailed t-test), but the two other differences are not statistically significant. It is impossible to conclude that the Internet setup somehow made the task easier and is responsible for the higher number of identified contexts, since the listener groups are not entirely comparable. The listeners in the Internet group were generally more experienced language users, and two are native speakers of English. In addition, part of the difference is caused by low scores from two of the listeners in the language lab group (18 correct contexts each), while the majority of listeners identified around 30 contexts in all three groups. All in all the differences between the groups are too small to be a major concern.

The differences as a possible result of different language backgrounds cannot be tested quantitatively, since there are relatively few listeners – between one and three – in each 'group', but their scores do not seem to deviate much from the overall result.

Of the 12 listeners with other than (just) Danish background two were excluded because of missing answers or failure to perform above chance level, leaving ten among the 56 listeners which have been analysed. One, with Romanian background, identified 24 of the 35 contexts, which is below average for the whole group, but the scores for the other nine were high. A Spanish listener identified 30 contexts; a Danish/Japanese listener identified 31; three Faroese listeners identified 29, 30 and 31 contexts respectively, and an Icelandic listener was one of only two listeners in the experiment who identified 33 contexts (the highest of any listener).

Finally, the three native speakers of English identified 30 (2) and 31 contexts, just above the overall average of 28.6 and very similar to the other non-Danish listeners. So in general the non-Danish listeners were among the ones with the highest number of identified contexts; the average (mean value) for this group is 29.9 compared with 28.3 for the Danish listeners, but the difference is not statistically significant.

7.5.3.2 Distribution of errors – response patterns

It is much more difficult to examine differences in the distribution of errors, that is, to find intra-group similarities and inter-group differences between the responses. The number of possible patterns is almost infinite – with three to five contexts for each of the 35 utterances the number is 1.29 raised to the power of 21 – so finding types of response patterns for 56 listeners, with the number of errors as a separate parameter, is very difficult. Graphs of expected and observed responses were produced for each listener in an attempt to visually determine recurring patterns, or listener profiles, but no clear patterns could be found. An example of such a graph can be seen in Figure 7.3.

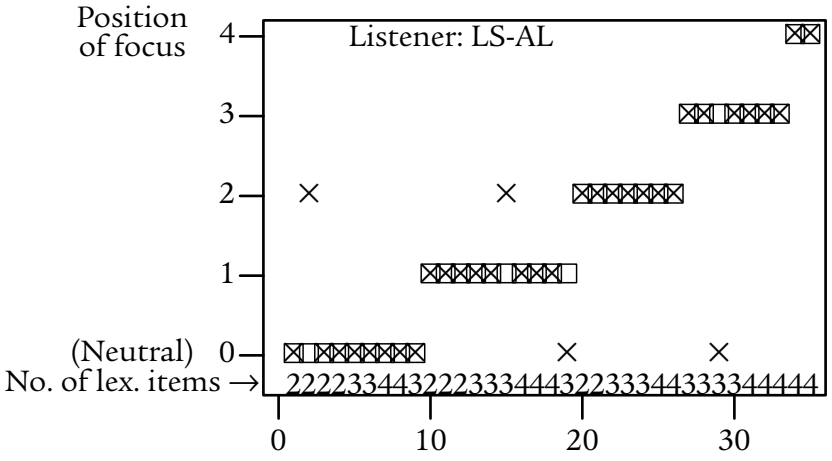


Figure 7.3. Expected/correct answers (□) and observed responses (×) for listener LS-AL.

It seems that finding patterns should be hypothesis-driven, that is, based on assumptions about where the differences might be. One such assumption which was examined in more detail was the tendency for listeners to perceive the utterance as

'neutral'. It must be expected that most errors (misidentified contexts) involve hearing a neutral context when an explicit focus was intended, or vice versa, rather than between two different locations of focus, and variation in the number of 'neutral'-responses may reflect a variation in sensitivity to the meaning, or function, of higher prominence on the final lexical item, that is, as default nucleus versus focal or contrastive accent.

The analysis did show that most of the misidentifications (328 of 360 total[3]) were between neutral versus focal interpretations, while only 32 were between different focal interpretations. The latter type of error was found both among high-scoring and low-scoring listeners, and with regard to the first type of error some listeners had a tendency to use the 'neutral' category where specific focus was intended, while others tended to perceive specific focus as neutral. Most, however, made errors in both directions and so it was not possible to find any clear patterns, or listener profiles.

Examining the responses of the three groups defined by the three test setups in the same manner did not reveal any systematic patterns or preferences for these groups. All further analyses were therefore performed on the pooled results of all 56 listeners.

7.5.4 Identified and misidentified contexts

The simplest way to present the data is to show the proportion (indicated here as the percentage) of the 56 listeners who perceived each of the possible contexts for each utterance, as in this example:

Utt	*Spk*	*Neut*	*Prominence + perceived focus*
pc f1	3F	5	The party 2.94 93 was cancelled 1.62 2

with information about intended context and speaker identity; the proportion of listeners (in per cent) who perceived a neutral context, and the proportion who heard explicit focus on each of the other lexical items in the sentence below these words.[4] The mean prominence ratings (English and Danish raters grouped) are printed above the words. The example shows that the sentence *pc*, produced by speaker 3F and intended as having focus on the first lexical item (*f1*) was perceived as neutral by 5% of the listeners (3 individuals). 93% of the listeners (52 individuals) heard focus on the first item, and 2% (1 listener) heard focus on the second, and last, item. The prominence ratings show that 'party' was clearly the most prominent word in the utterance, with a score of close to 3 (strong stress).

[3] The total number of judgements in the test is 1,680 (56 listeners × 30 items), yielding 1,320 correct identifications. The error rate is 21.4%.

[4] 'Hearing focus' on an item is used here and in the following as a shorthand expression of 'assigning the context which in the recordings was designed to produce focus on this item'.

These are the results for all 35 utterances, sorted according to intended context/focus structure:

Utt	*Spk*	*Neut*	*Perceived focus + prominence*
ps n	4M	71	2.06 2 Paul sings 14 14
ps n	5M	39	1.94 2.25 Paul sings 2 59
pc n	3F	54	2.62 2 The party was cancelled 43 4
pc n	6M	50	2.12 2.31 The party was cancelled 7 43
bsa n	3F	16	2.81 1 2.19 Bill struck Ann 79 5
css n	2F	82	2.06 1.94 2 The cook was smelling the soup 11 4 4
jkft n	5M	91	2 1.69 1.81 2.25 Jane kissed Frank tenderly 9
sepc n	1F	79	2 1.94 1.88 2 Sheila examined the patient carefully 18 4
pdp n	2F	66	2.12 1.75 2.19 Is Peter a doctor in Paris 34
ps f1	4M	0	2.94 1.38 Paul sings 100
ps f1	5M	2	2.88 1.31 Paul sings 98
pc f1	3F	5	2.94 1.62 The party was cancelled 93 2
bsa f1	3F	0	2.88 1.12 1.75 Bill struck Ann 100
css f1	2F	0	2.75 1.44 1.38 The cook was smelling the soup 100
css f1	5M	2	2.94 1.38 1.50 The cook was smelling the soup 91 7
jkft f1	1F	25	2.69 0.94 1.62 1.69 Jane kissed Frank tenderly 68 2 5
jkft f1	5M	0	2.94 1.06 1.50 1.62 Jane kissed Frank tenderly 100

Utt	*Spk*	*Neut*	*Perceived focus + prominence*
sepc f1	1F	2	2.94 1.56 1.62 1.69 Sheila examined the patient carefully 96 2
pdp f1	2F	27	3 1.50 1.88 Is Peter a doctor in Paris 73
ps f2	1F	20	1.69 2.56 Paul sings 80
pc f2	2F	16	1.75 2.81 The party was cancelled 2 82
bsa f2	1F	0	1.75 3 1.94 Bill struck Ann 98 2
bsa f2	4M	0	1.62 3 1.50 Bill struck Ann 2 98
css f2	6M	7	1.44 3 1.38 The cook was smelling the soup 2 91
jkft f2	4M	4	1.69 3 1.38 1.50 Jane kissed Frank tenderly 95 2
sepc f2	6M	4	1.31 3 1.44 1.44 Sheila examined the patient carefully 93 4
bsa f3	4M	4	2 1.06 2.81 Bill struck Ann 2 95
bsa f3	6M	4	1.50 0.88 3 Bill struck Ann 2 95
css f3	2F	63	2 1.81 2.31 The cook was smelling the soup 9 29
css f3	3F	9	1.88 1.56 3 The cook was smelling the soup 2 89
jkft f3	3F	4	1.94 1.38 3 1.31 Jane kissed Frank tenderly 2 93 2
jkft f3	6M	5	1.75 0.75 3 2.12 Jane kissed Frank tenderly 4 89 2
sepc f3	4M	0	1.75 1.31 3 1.19 Sheila examined the patient carefully 2 98
jkft f4	1F	34	1.94 1.44 1.81 2.38 Jane kissed Frank tenderly 66
jkft f4	3F	0	1.88 1 1.44 2.75 Jane kissed Frank tenderly 2 98

The reason for indicating both the proportion of scores for each context and the prominence ratings from the previous experiments is of course the predicted strong association between prominence and perceived information structure. This association is sometimes obvious in the data and sometimes less obvious.

7.5.4.1 Utterances with a specific focus

There are 26 utterances with an intended specific focus (*f1-4*). In 22 of these agreement among the listeners is 75%[5] (42 of 56) or higher, which must be considered quite good agreement. In all of these the focused item was deemed very prominent, almost always between 2.75 and 3 on the scale from 0 to 3, and the non-focused items often had low, or reduced, prominence. Figure 7.4 shows some examples of typical cases.

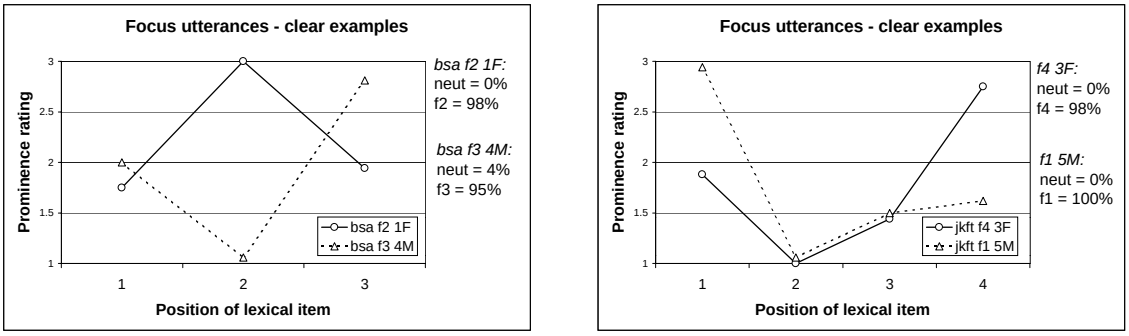


Figure 7.4. Examples of utterances with good listener recognition of the intended focus. The prominence ratings are from Tests 1 and 2 (grouped). The per cent scores indicate the proportion of identifications for that context.

The final, post-focal item in utterance *bsa f2* 1F is deemed more prominent (1.94) than the corresponding item in most of the other utterances, where the prominence ratings typically vary between 1.3 and 1.7. This is not reflected in the identification of the contexts, though. All listeners except one identified the context correctly. The connection between perceived prominence and perception of information structure is very clear in these examples, and in most of the other utterances where agreement is high, through a combination of high prominence on the focused item and (often) reduction on the non-focal items.

The four utterances where agreement was somewhat lower are not uniform in their deviation from the general pattern. In utterance *pdp f1* by speaker 2F the focused item is very prominent (3.0) and the two non-focused items are both reduced from normal stress, but over 25% of the listeners still perceived the utterance

[5] The percentages used in this section denote the proportion of listeners who responded in a particular way. It is thereby different from the pairwise per cent scores used in Tests 1–3, which denoted the proportion of pairwise agreements.

as neutral. The reason for this may be that the task was slightly different for this utterance: focus was not elicited by asking a question but by providing a (possible) context. This may have produced some confusion and hence uncertainty about which of the two possible answers was correct. In addition, this utterance was produced with a falling-rising tune, that is, a fall on 'Peter' and a rise on 'Paris'. This too may have added to the difficulty of the task, considering the uncertainty about the prominence level of such final rises which was noted in the previous experiments.

Two utterances – *css f3* 2F and *jkft f4* 1F – are very similar in terms of prominence levels. The final lexical item is focused but scored only 2.3–2.4 on the prominence scale, which is closer to normal stress than to strong stress. It is therefore not surprising that many listeners heard the utterances as neutral. Figure 7.5 shows the prominence ratings and proportions of identified contexts of both utterances. It is not immediately clear from the prominence ratings why so many more listeners mistook utterance *css f3* 2F for a neutral utterance. The apparent prominence reduction on pre-focal items in utterance *jkft f4* 1F is not much different from what is found in many neutral versions of this utterance.

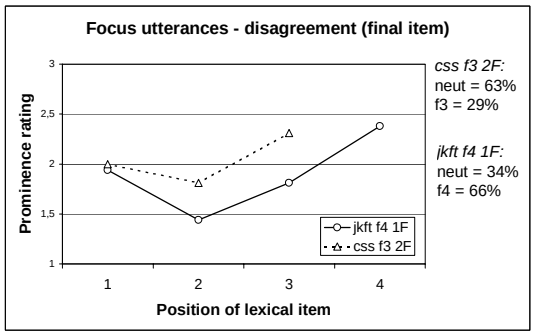


Figure 7.5. Two utterances with large disagreement about the contexts. Prominence is fairly low on the intended (final) focal item.

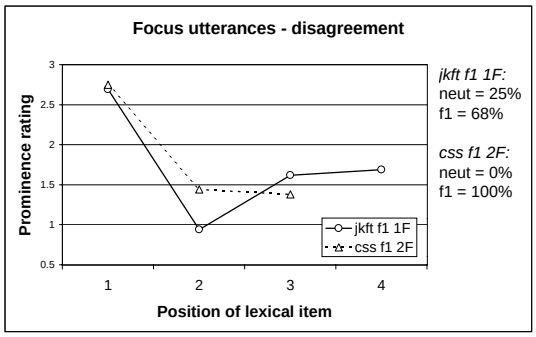


Figure 7.6. Utterance *jkft f1* 1F has a higher proportion of 'neutral'-responses than expected from prominence ratings (compared with *css f1* 2F).

Utterance *jkft f1* 1F was perceived as neutral by 25% of the listeners, which is not to be expected from the prominence ratings. Although the prominence level of 2.69 on the focused item is not among the highest, it is not much lower than in utterance *css f1* 2F where all 56 listeners identified the context. The utterances are depicted in Figure 7.6 which shows that both have fairly strong perceived prominence on the focal item and reduction on post-focal items. Whatever made 25% of the listeners perceive utterance *jkft f1* 1F as neutral does not seem to be directly linked to prominence relations.

7.5.4.2 Neutral, context-free utterances

Agreement is generally lower about the neutral utterances; it is higher than 75% in only three out of nine cases. In one of these – *jkft n* 5M – the final lexical item is slightly more prominent (2.25) than the other items, but in the other two – *sepc n* 1F and *css n* 2F – it is as prominent as the first item or even, in the latter case, slightly less prominent. Utterance *ps n* 4M is similar to *css n* 2F in terms of perceived prominence, and 71% of the listeners agree that it is neutral, while the remaining 29% are equally divided between hearing focus on either of the two lexical items. These observations seem to confirm the general pattern of prominence relations in neutral utterances, which was established in Tests 1 and 2. The first and last lexical items are generally more prominent than any other items, and the last item can be as prominent, slightly more prominent or less prominent than the first. If these conditions are met the utterance will be perceived as neutral.

Some of the other neutral utterances in this experiment contradict these expectations, however. Utterances *ps n* 5M and *pc n* 6M with two lexical items have slightly more prominent final items, and seem similar to *jkft n* 5M. But 59% and 43% of the listeners, respectively, heard focus on the final item. In *ps n* 5M this might be caused by the slight reduction in prominence on the first item (the value is 1.94), although the reduction is too small to make this explanation an obvious one. In utterance *pc n* 6M the first item is not reduced (the value is 2.13) and the difference between the two lexical items is quite small. The prominence relations can therefore not easily explain why 43% of the listeners heard a focus on the final item, while only 9% heard the final item of *jkft n* 5M as focused.

Utterance *jkft n* 5M can be compared with utterance *jkft f4* 1F, which was also commented on above and depicted in Figure 7.5. These two utterances have very similar prominence relations, as can be seen from Figure 7.7.

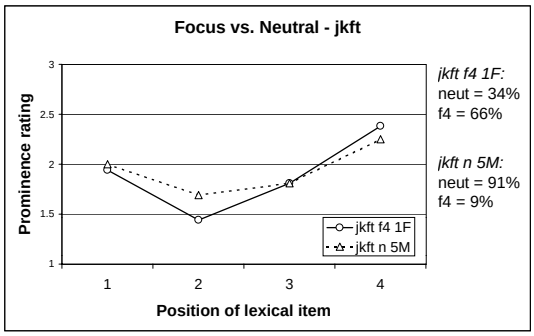


Figure 7.7. Focused (*f4*) and neutral version of *jkft*. The prominence ratings are similar, but their contexts were perceived as different.

Although 25% of the listeners heard *jkft f4* 1F as neutral that is still much below the 91% who perceived *jkft n* 5M as neutral – a difference which it is difficult to relate to the small differences in prominence ratings. If the connection is to be found in the prominence ratings the question arises whether it is caused by the slightly more prominent final item in *jkft f4* 1F or the slightly less prominent second item (pre-

focal reduction), or (as is most likely) a combination.

Two utterances deviate from the expected pattern in a different way: *pc n* and *bsa n* by speaker 3F were two of the utterances which were excluded from analysis in Tests 1–3, because of the strong prominence on the first lexical item (2.62 and 2.81 respectively). The anomaly of the speaking style of this speaker is confirmed by the present results, since many listeners heard a focus on the first item (79% and 43% respectively). But it is perhaps surprising that these numbers are not even higher, particularly for *pc n* 3F, considering that one item seems to have stood out so clearly above the other. The corresponding versions with focus on the first item were also included in the experiment; the scores of all four utterances are depicted graphically in Figure 7.8.

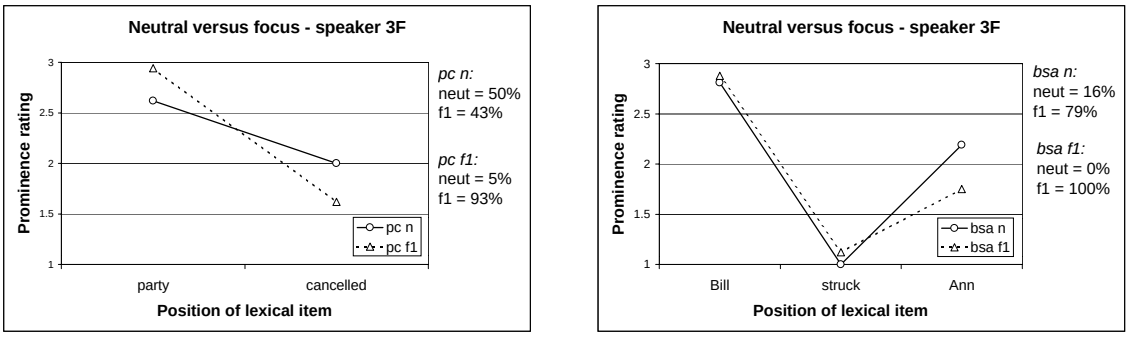


Figure 7.8. Sentences *pc* (left) and *bsa* (right) by speaker 3F in neutral versions (n) and versions with focus on the first lexical item (*f1*).

The two versions of *bsa* are very similar in terms of prominence, the only difference being that the neutral version has a more prominent final item, while the first item is very prominent in both utterances. This suggests that it is the lack of reduction of the final item which is designed to signal the neutral context, or broad focus, in this particular case. This signal is not successfully decoded by most of the listeners in this experiment: 79% heard focus on the first item of the neutral version. Still, it is interesting that 16% of the listeners did hear the utterance as neutral despite the fact that the first item stands out so clearly above the rest. It may be an indication of the importance of stress reduction of post-focal items for the perception of information structure.

The focused (*f1*) version of *pc* by the same speaker displays both higher prominence on the focused item than the neutral version and reduction on the single post-focal item and is consistently recognised as having a specific focus, just like *bsa f1*. But here the neutral version makes an even more striking comparison. This utterance has a very prominent first lexical item (2.62) and a less prominent second, and final, item (2.0). Yet this utterance is recognised as neutral by 50% of the listeners, which is better than random agreement (χ^2 goodness-of-fit test). The explanation does not seem to lie in the prominence of the first element, which is only slightly

lower than in utterance *css f1* 2F (see Figure 7.6) where all listeners heard focus on the first item. Rather, it might be the prominence of the second item, which although much less prominent than the first is still 'unreduced', that is, it was perceived as having normal, full stress.

The (relatively few) observations of regular patterns and apparent deviations in this experiment suggest the following explanation of the connection between perceived prominence and the perception of information structure.

- If the lexical items (especially the first and last items) are equally prominent and generally perceived as having 'normal, full stress' the utterance will be perceived as neutral, or context-free.
- If one item is (much) more prominent than all other items, it will be perceived as focused. However, there is an asymmetry: a larger difference is required between a focused item and any post-focal items to signal focus than between the focused item and pre-focal items. This is particularly clear with utterance-initial and utterance-final focus.
- The connection between perceived prominence and information structure is most clearly expressed by the degree of prominence of the most prominent item, but reduction of non-focal items (especially in post-focal position, see above), or rather the relation between the focused and non-focused item(s) seems to contribute, and may under certain conditions act as an important cue.

I have illustrated some of these principles in Figure 7.9, using the simplest case of a two-stress utterance as an example.

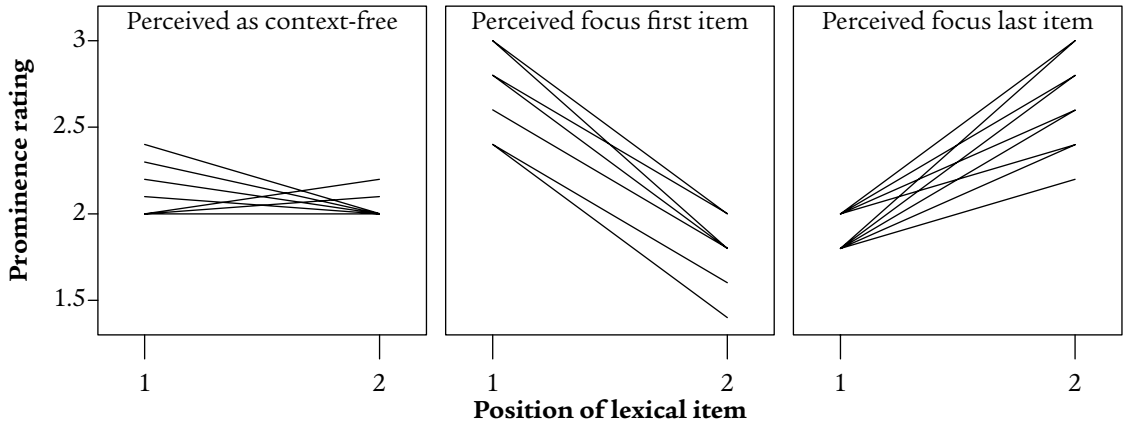


Figure 7.9. Illustration of the connection between perceived prominence and the perception of information structure. The lines indicate typical prominence relations between the initial and final lexical item in a two-item utterance.

In utterances with more than two lexical items the relative contributions of the non-focal items cannot easily be determined from the present data, but I propose the following hypotheses based on the prominence ratings in Tests 1–3 and the few

relevant examples in the present experiment.

- Reduction of the prominence of non-focal items will promote a focus-interpretation, and normally, all non-focal items will be reduced. If only some items are reduced it may lead to the perception of two informational foci, for example 'JANE kissed Frank TENderly' (in response to a putative question 'Who kissed Frank in what way?'). No provision was made in the experiment for this type of interpretation.
- All post-focal items must be reduced. Otherwise the utterance is likely to be interpreted as neutral (or as having two foci).
- The second, but non-final, lexical item is often reduced even in neutral utterances, and extra reduced in utterances with a focus on the first item. Subsequent items may be more prominent than this item without being heard as focused, as long as they are reduced from normal stress.

These hypotheses could be tested by repeating the experiment and including utterances with systematically different prominence levels on the non-focal items. It would also be interesting to allow the listeners to indicate two (or more) foci.

The results show that there is a very strong association between perceived prominence and the perception of information structure (as expected), but that the relation is not always straightforward. In the next section I will explore a few simple hypotheses about the relation and see to what degree the perceived information structure can be inferred from the prominence relations in an utterance.

7.6 Inferring information structure from prominence relations

Ideally, the objective is to establish a complete map of the relation between prominence and information structure, that is, to be able to predict the proportion of listeners who would perceive focus on a particular word, or no focus (context-free utterance) given the prominence ratings above the words in the presentation of data at the beginning of Section 7.5.4. However, that would require some fairly complex statistical procedures, and the material is not comprehensive enough to justify these. Instead, some of the hypotheses formulated in the previous section are tested, using simple correlation and regression analysis. The question which is explored is simple: given the prominence ratings what will be the proportion of perceived 'neutral, context-free' answers? It was reported earlier that less than 10% of the misidentifications concerned the *location* of focus, that is, where a listener heard focus on a different item than the intended. In general, if an utterance was heard as having (narrow) focus, the location of this focus was also identified. The main problem is therefore reduced to whether there is a specific focus or not. If yes, it can be assumed that the most prominent item in the utterance will be heard as focused. The dependent variable of the regression analyses will therefore be the number of 'neutral'-responses, as an expression of the likelihood that the utterance will be perceived as context-free (if it is high) or as having a specific focus (if it is low).

7.6.1 Hypothesis 1

Hypothesis 1 can be stated as follows:

> The probability that an utterance will be perceived as having a specific focus is proportional to the prominence rating of the most prominent item in the utterance.

This is the simplest possible statement of the relation. It makes the implicit assumption that focus is only indicated locally, on the focused item itself, without any contribution from surrounding items, or that the contribution from other sources is predictable from (that is, covaries with) the most prominent item. The prominence value of the most prominent item will be referred to in the following (both text and graphic displays) as *MaxProm*. The result of the regression analysis is presented in Table 7.2 and Figure 7.10.

Correlation and regression statistics			
r	0.913	*S.e.*	11.844
r^2	0.833	*N*	35
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	234.414	16.665	0.000
MaxProm	-78.072	6.091	0.000
s.e. = standard error, here and in the following			

Table 7.2. Correlation and regression statistics for hypothesis 1 – information structure predicted from MaxProm.

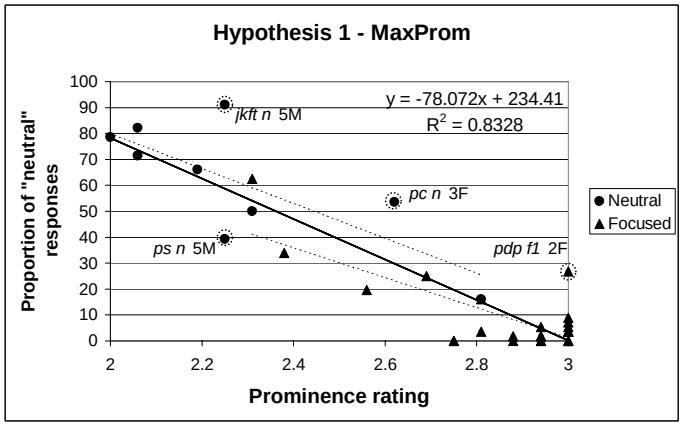


Figure 7.10. Plot of the regression analysis of the relation between the level of prominence on the most prominent item in an utterance (MaxProm) and the number of raters who perceived the utterance as context-free. Neutral and focused utterances are shown with separate symbols and regression lines (dashed). The formula is for the regression line for all utterances (solid).

The correlation between MaxProm and perceived focus structure is very strong (r = 0.913) and can account for 83% of the variance in the material. The simplest possible hypothesis therefore seems to be a very powerful one.

The intended neutral and focused utterances are generally grouped at either end of the scale, as expected, with some overlap. For both groups it is clear that variation in the prominence level of the most prominent item also leads to variation in perceived structure. The 'outliers' which are not well accounted for by the simple model are enclosed in dotted circles. In the neutral group they comprise three out of the nine utterances. Two – *ps n* 5M and *pc n* 3F – caused a large degree of disagreement among the raters and were commented on in Section 7.5.4, while agreement was high about utterance *jkft n* 5M. It is particularly difficult to explain why the two utterances by the same speaker, with similar perceived prominence, are perceived so differently in terms of information structure.

In the group of utterances with intended focus the clearest outlier is utterance *pdp f1* 2F which, as was argued in Section 7.5.4, is probably not related to phonetic cues but to the use of a different kind of context (necessitated by the fact that it is an interrogative).

7.6.2 Hypothesis 2

Hypothesis 2 concerns the prominence of non-focal words:

> Reduction of the prominence level of non-focal items contributes to the perception of focus structure and may in some cases be an important cue.

This hypothesis addresses the role that backgrounding the non-focal items plays for the perception of information structure. The issue is very complex because factors such as the position of the non-focal item (pre- or post-focal) and distance from the focal item may have a significant effect, but here only a relatively simple expression of the effect of reduction is tested.

For each utterance the mean prominence level of items which are not the most prominent is calculated, as in this example:

jkft f3 $^{1.83}$Jane $^{0.79}$kissed $^{2.92}$Frank $^{0.92}$tenderly

The most prominent item (here *Frank*, at 2.92) is used as one independent variable. The mean prominence value of the other items (here 1.62) is used as the second independent variable. It will be referred to in text and figures as *NonfocMean*.

A simple regression analysis was also performed, using the difference between MaxProm and NonfocMean as the independent variable. This captures most of the same information and the result can be plotted in a (normal) two-dimensional diagram, Figure 7.11. The results of the regression tests are shown in Table 7.3.

Multiple regression			
R	0.927	*S.e.*	11.066
R^2	0.858	*N*	35
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	143.711	40.74	0.001
MaxProm	-62.305	8.67	0.000
NonfocMean	29.273	12.15	0.022

Simple regression			
r	0.920	*S.e.*	11.373
r^2	0.846	*N*	35
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	75.22	4.375	0.000
MaxProm – NonfocMean	-48.92	3.636	0.000

Table 7.3. Regression statistics for hypothesis 2 (contribution of less prominent items).

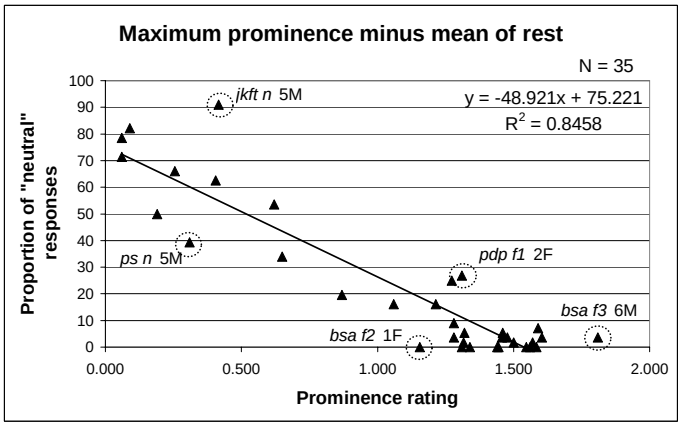


Figure 7.11. Simple regression test of the difference between the highest prominence level in an utterance (MaxProm) and the mean value of the remaining prominence values (NonfocMean). Table 7.3.

The correlation is marginally better than for hypothesis 1, and the contribution of the second independent variable is significant but quite small. Although this may indicate an effect of non-focal reduction, it is hardly strong evidence in favour of hypothesis 2.

The outliers according to this model are mostly the same as according to hypothesis 1: *pdp f1* 2F still seems somewhat deviant, and it is still unclear why utterances *ps n* and *jkft n* by speaker 5M are perceived so differently. Utterance *bsa f2* 1F received fewer 'neutral'-responses than the model predicts, and this is in agreement with the comments made about this utterance at the beginning of Section 7.5.4.1, that the final item is slightly more prominent (at 1.94) than would be expected in

post-focal position. Utterance *bsa f3* 6M is deviant simply because the prominence scores lead to a *very* low predicted proportion of 'neutral'-responses but two listeners still heard it as neutral, so this can be seen as a kind of ceiling effect – the model is simply not accurate at the extreme ends of the scale. Utterance *pc n* 2F is better explained by this model, as predicted in Section 7.5.4.

7.6.3 Hypothesis 3

Hypothesis 3 relates to the relative contribution of pre- and post-focal prominence levels:

> Post-focal reduction is more important for the perception of focus than pre-focal reduction.

This hypothesis concerns the observation from Tests 1 and 2 that post-focal items were generally less prominent, or more reduced, than pre-focal items, and the, admittedly spurious, cases in this experiment where the prominence level of a post-focal item was suggested as a possible explanation of listener responses.

One of the assumptions is that a non-reduced post-focal item will block the interpretation of this item (and perhaps a larger domain) as backgrounded and thereby affect the perception of focus in the utterance. The same may happen in pre-focal position, but as the hypothesis states, to a lesser degree. This is also related to the observations that no pitch excursions are normally found in post-focal position (e.g. Nakatani and Aston 1978), and that lexical stress distinctions are neutralised in this position (Huss 1978). The options available for marking some degree of stress or prominence simply seem more limited post-focally, perhaps to accommodate the need for unambiguous signalling of information structure.

The prominence ratings of the most prominent pre- and post-focal items are used as an expression of pre- and post-focal reduction respectively. They are referred to in the following as PrefocMax and PostfocMax. In simple regression tests both these variables were found to correlate positively with the proportion of perceived 'neutral'-responses: the lower the prominence value, the lower the proportion of perceived 'neutral'-responses (PrefocMax: $r = 0.628$, s.e. = 23.323, $p < 0.001$; PostfocMax: $r = 0.600$, s.e. = 22.164, $p < 0.001$). Although the confidence level of both coefficients is significant, they are not very high and much lower than the coefficients for the variable MaxProm. The interesting question is whether the pre- and post-focal contexts can contribute *additional* information and improve the concerted predictive power of the equation. Both variables (PrefocMax and PostfocMax) were therefore inserted in regression analyses with two independent variables, the other variable being MaxProm in both cases. As in the previous section, the 'focal accent' is here defined as the most prominent word in the utterance. This definition allows both the context-free and the focused utterances in the test to be used. Since the correlations can only be calculated for sub-groups of the material, the simple test (hypothesis 1) for the same data sets is also included for comparison.

Correlation and regression statistics			
Pre-focal			
Multiple R	0.919	*S.e.*	12.146
R^2	0.844	*N*	21
Simple r	0.918	*S.e.*	11.830
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	236.082	55.331	0.000
MaxProm	-76.965	10.667	0.000
PrefocMax	-2.720	17.247	0.876
Post-focal			
Multiple R	0.947	*S.e.*	9.129
R^2	0.897	*N*	23
Simple r	0.935	*S.e.*	9.831
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	191.199	28.409	0.000
MaxProm	-71.698	7.037	0.000
PostfocMax	15.912	7.625	0.049

Table 7.4. The contribution of pre- and post-focal reduction separately. Coefficients for the simple regression analysis of the same utterances are included for comparison

As can be seen in Table 7.4, the contribution of pre-focal reduction is non-significant (p = 0.876) and adding this variable did not improve the correlation between the prominence ratings and the number of perceived neutral contexts. Post-focal reduction had a minor influence on the correlation, and the contribution from this variable is (only just) significant at the 5% level. This indicates a difference between pre- and post-focal reduction as predicted by the hypothesis, although the difference is fairly small.

The relative contribution of the two relatively weaker predictor variables PrefocMax and PostfocMax can be examined (also graphically) more closely by keeping the stronger independent variable – MaxProm – constant and then calculating predicted values for the variable we are interested in. That is, to examine the contribution of the variable PrefocMax we can calculate predicted values using this formula:

$$y' = a + b_1 MaxProm_{mean} + b_2 PrefocMax$$

y' = predicted value
a = Y-intersection
b_1 = regression coefficient MaxProm
b_2 = regression coefficient PrefocMax

The variable MaxProm is kept constant; in this case by using the arithmetic mean of

all values. The procedure for the other variable (PostfocMax) is the same, only using a constant for PrefocMax instead. When the new predicted values are plotted (on the y-axis) against the dependent variable (proportion of 'neutral'-responses) on the x-axis they will form a straight line. This line represents the contribution of the variable when the other (dominating) value is kept constant. The contributions of pre- and post-focal prominence are shown in Figure 7.12 and Figure 7.13.

The regression line for the pre-focal condition shows almost no variation with prominence rating. To the extent that it does vary the correlation is negative, that is, the opposite of what was predicted (but non-significantly so). The regression line for the post-focal condition varies positively with prominence rating, which means that in addition to the correlation with MaxProm there is a slight effect of post-focal prominence: when the level of post-focal prominence increases, listeners are less likely to perceive the utterance as neutral. This is more or less as expected, although the complete lack of an effect of PrefocMax was perhaps not as predicted.

If post-focal (and pre-focal) reduction act as a significant and independent cues to the perception of information structure we should expect a more pronounced

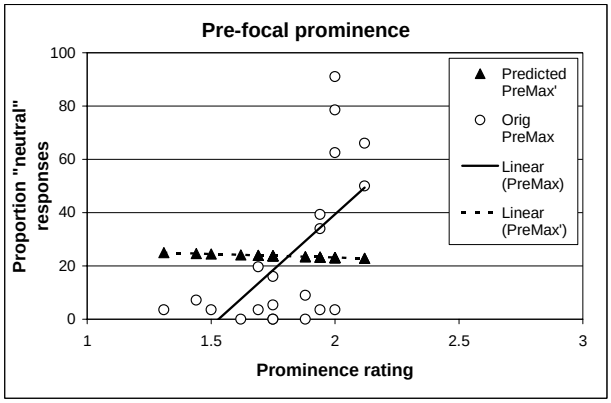


Figure 7.12. The contribution of pre-focal maximum prominence (PrefocMax) when the maximum (peak) prominence in the utterance (MaxProm) is kept constant.

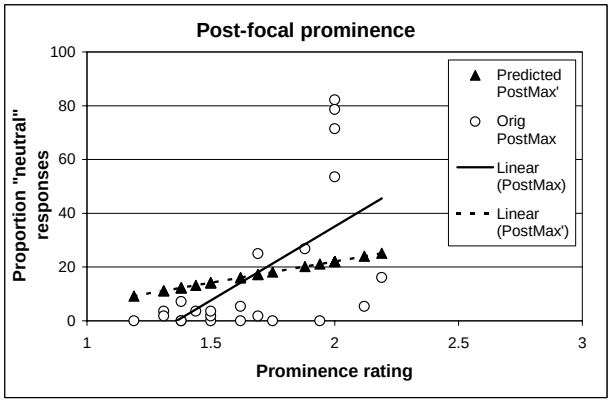


Figure 7.13. The contribution of PostfocMax when MaxProm is kept constant.

effect. But one of the problems may be that this phenomenon is not independent but covaries with the level of prominence of the most prominent accent, in which case it is difficult to capture their separate contributions in a regression test such as above, or even impossible in the case of absolute covariation. This is relevant to the results of hypotheses 2 and 3, and the correlations between MaxProm and the tested variables in these tests are shown in Table 7.5.

Correlation and regression statistics			
NonfocMean – MaxProm			
r	0.755	*S.e.*	0.222
r^2	0.569	*N*	35
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	4.445	0.264	0.000
NonfocMean	-1.057	0.160	0.000
PrefocMax – MaxProm			
r	0.695	*S.e.*	0.261
r^2	0.483	*N*	21
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	4.733	0.487	0.000
PrefocMax	-1.124	0.267	0.000
PostfocMax – MaxProm			
r	0.503	*S.e.*	0.283
r^2	0.253	*N*	23
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	3.712	0.346	0.000
PostfocMax	-0.545	0.204	0.014

Table 7.5. Covariation between MaxProm and the level of non-focal prominence as expressed in hypotheses 2 and 3 (and used as independent variables in the multiple regression analyses).

The correlation between the two variables (or possible cues to information structure) is statistically significant in all three cases, but the correlation coefficients are not particularly high, especially for the post-focal context. The difference between pre- and post-focal context may be partial explanation of why there was no significant effect of adding pre-focal context in hypothesis 3, while there was slight effect of adding post-focal context, but the results are not conclusive.

Another possible effect which might be considered is interaction between the cues. It is likely that pre- or post-focal reduction is a more efficient cue when other cues are less clear. MaxProm was referred to as a *dominating* cue above, because of its significantly larger explanatory power, and one obvious hypothesis could be that if

this cue is strong the interpretation of information structure becomes unambiguous, but when it is less strong other cues may influence perception. In other words, it is worth examining the interaction between MaxProm and the non-focal reduction represented by PrefocMax and PostfocMax. This interaction can be captured by multiplying MaxProm with the variable we are interested in and including this product in a three-way multiple regression analysis with the simplex variables as the other independent variables. The results for the two tests (PrefocMax and PostfocMax) are shown in Table 7.6.

Table 7.6. Interaction effect between the strongest predictor variable MaxProm and the two variables PrefocMax and PostfocMax. The Interaction variable is the product of MaxProm and Pre- or PostfocMax.

Multiple regression – interaction effects			
Interaction between MaxProm and PrefocMax			
R	0.934	*S.e.*	11.346
R^2	0.872	*N*	21
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	-420.221	348.446	0.244
MaxProm	148.158	118.620	0.229
PrefocMax	328.605	174.706	0.077
Interaction	-113.991	59.851	0.074
Interaction between MaxProm and PostfocMax			
R	0.959	*S.e.*	8.255
R^2	0.920	*N*	23
Variable	*Parameter*	*S.e.*	*p*
Y-intercept	-458.577	279.313	0.117
MaxProm	149.919	95.074	0.131
PostfocMax	345.304	141.161	0.023
Interaction	-112.492	48.151	0.031

Table 7.6 shows some interesting results, both for the pre-focal and for the post-focal condition. The interaction effect is significant for the post-focal condition (PostfocMax, $p = 0.023$) and the significance level of the PostfocMax variable itself is higher than in the previous test. In other words, the efficiency of post-focal reduction as a cue to information structure depends on the level of the prominence peak (MaxProm) in the utterance. This effect is plotted using different constant values of MaxProm in Figure 7.14 b). Notice also that the correlation coefficients are very high: 92% of the variance can be explained from the variables MaxProm, PostfocMax and the interaction between them. But even for the pre-focal condition there is an interesting effect of including interaction. Although the interaction is not significant ($p = 0.074$) it is quite strong, and just as importantly, the effect on the variable PrefocMax

of keeping MaxProm constant is quite large: the significance level increases from 0.876 to 0.077, so although it just fails to reach significance at the 5% level there is an appreciable effect of adding interaction as a variable. It should also be noted that there were only 21 observations in the test (PrefocMax); it is not unreasonable to assume that the effect would have been significant with more observations. The effect of interaction between MaxProm and PrefocMax is displayed in Figure 7.14a.

The diagrams in Figure 7.14 are produced in the following manner: predicted values including the interaction effect are calculated using the formula:

$$y' = a + b_1x_1 + b_2x_2 + b_3x_1x_2$$

where
a = Y-intercept
b_1 = slope of MaxProm
b_2 = slope of Pre- or PostfocMax respectively
b_3 = slope of Interaction
x_1 = constant MaxProm value
x_2 = Pre- or PostfocMax variable respectively

Different values were used as a constant for MaxProm to capture the observed variation in the dataset, namely 2 (normal, full stress), representing the lower limit; 3 (strong stress), representing the absolute upper limit; the mean value of the tested variable; and (chosen post-hoc) 2.4, which is intermediate between the lower limit 2 and the mean value.

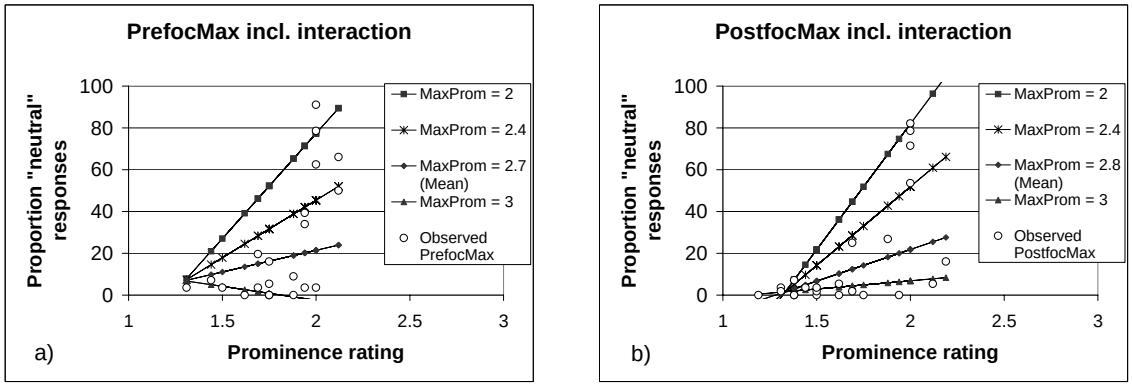


Figure 7.14. The four lines represent predicted correlation between prominence ratings and proportion of 'neutral'-responses at different (fixed) levels of MaxProm when interaction with PrefocMax (a) or PostfocMax (b) is considered. Observed values are included for reference.

The diagrams in Figure 7.14 show the different effect of the pre- and post-focal maximum prominence on perceived information structure as a function of different values of MaxProm. When MaxProm is 3 then the contribution of PrefocMax is negative and that of PostfocMax is quite small, but as MaxProm decreases the effect of

the other variables increase. This effect was illustrated by utterance *bsa f2* 1F (see Section 7.5.4.1), which was not perceived as neutral by any listeners despite the relatively prominent final item (1.94): the effect of the most prominent item (a prominence value of 3) superseded the influence of the post-focal item.

Another factor may have influenced the relative contributions of PrefocMax and PostfocMax in these tests, namely the larger variation in PrefocMax values. This variation is not just correlated with MaxProm (as shown above) but also with the position of the focal accent in the utterance: pre-focal items are reduced in inverse proportion to their distance from the focal item (see Section 4.3.2). This systematic variation has not been included in the tests, and it may be that pre-focal reduction could be shown to correlate as highly with perceived information structure as post-focal reduction if this parameter were included, but the data set is too small for the required analyses. In any case, it does not affect the conclusion from Tests 1–3 that pre- and post-focal reduction behave differently and that post-focal reduction is more pronounced.

In general, while the statistical tests did lend some support to the hypotheses which were formulated on the basis of the qualitative analysis in Section 7.5.4, the results were not conclusive. In addition to the problem of partial covariation between several of the tested variables, it also seems clear that a solid quantitative analysis would require more data. The possible influences on the perception of information structure are numerous – even if restricted to prominence relations – and a larger corpus would provide more precise information about this variation and whether the outliers in this experiment are inexplicable deviations or systematically different realisations.

7.7 Conclusion

Although the (partial) results were not conclusive, the answers that were posed at the beginning of Section 7.1 can be answered satisfactorily.

Re 1) The association between variations in prominence relations and perceived information structure is very strong. Over 83% of the variance in perceived information structure can be explained simply by reference to the prominence level of the most prominent item in the utterance. There is therefore cause to believe that the acoustic cues which prompted a specific perception of prominence will also provide information about the perception of focus relations in an utterance. This is perhaps not surprising, but is good validation of the results from Tests 1 and 2 as reflecting linguistically relevant and significant information.

Re 2) Listeners could and did distinguish clearly and consistently between neutral utterances in which the final item is more prominent than the others, and utterances where the final element is highlighted or focused. However, the prominence level of this item determines whether the utterance will be

perceived as neutral or focused (cf. (1) above) regardless of the intended focus structure of the utterance. There were one or two exceptions to this general pattern which could not be explained from the prominence relations.

Re 3) Again, following the general principle outlined under (1), non-final (here always initial) items which are perceived as the most prominent ones in neutral utterances may be perceived as highlighted. But the difference between the initial item and following items needs to be fairly large; in fact, a larger difference seemed to be required between (initial) focal and post-focal items than between focal and pre-focal items. This difference was attributed to a general stronger requirement of post-focal reduction, although the results of the quantitative tests were not conclusive.

There were several interesting deviations from the general pattern, but the material is not comprehensive enough to determine if these are systematic deviations which could provide more information about relevant cues to focus structure or simple statistical outliers. This could be tested by repeating the experiment with more utterances, especially ones where the prominence relations are counter-intuitive or just inconclusive.

Summary and future research

Summary

The experiments reported in this thesis demonstrated some regularities and trends in the manifestation of stress as perceived prominence:

- In neutral, context-free utterances the first and last stressed syllables, located in the first and last lexical words, were generally perceived as the most prominent ones. This is in accordance with most descriptions of English intonation in the traditional British framework, where these positions are often referred to as *onset* (or *head*) and *nucleus* respectively. In contrast to these descriptions and most other descriptions of English intonation, the last item (the nucleus) could not be shown to generally be the (single) most prominent syllable in the utterance (or intonation phrase), but appeared to be so in only 15–20% of the neutral utterances.
- Stressed words between the first and and last lexical items had reduced prominence in a strong – weak alternating pattern, although the reduction only very rarely amounted to what one might call complete deaccentuation.
- In utterances where one item was emphasised for semantic or contrastive focus it appeared that placing an item in such narrow focus had both a local effect – much higher perceived prominence of the focal accent – and a global effect – a general reduction of the perceived prominence of the other lexical items. In pre-focal position the level of reduction seemed to be associated with both (a) the absolute position of the item in the utterance (as first, second or third lexical item) and (b) the position relative to the focal accent: stresses were reduced in inverse proportion to their distance from the focal accent. In post-focal position only factor (a) seemed to be operational; all post-focal stressed words were reduced to (well) below 'normal, full stress' (as reported by listeners).
- The effectiveness of both local and global prominence levels as cues to information structure was demonstrated in a separate experiment, which also showed that when the final lexical item (the default location of a nucleus) is perceptually more prominent than (all) previous items, it is generally perceived as being in narrow focus. This can be interpreted as an argument against the expectation that nuclei in neutral utterances (in broad focus) should necessarily be made more prominent than other stressed syllables.
- A brief characterisation of the acoustic parameters F_0 and duration pointed to F_0 as a close and very direct correlate of perceived prominence. Both the local and

global effects of narrow focus on perceived prominence were mirrored in the F_0 traces: F_0 is 'boosted' on the focal accent, and in pre-focal position F_0 movements are reduced but clearly present, while F_0 movements are largely absent in post-focal position.

Future research

The observations above leave some questions open for future research. With regard to perceived prominence some of the obvious questions might be:

- Are the regular variations in perceived prominence level such as the strong – weak alternating pattern also present in other types of speech – most notably spontaneous dialogue – or is that an artefact of the speech situation of reading short, rhythmically relatively regular, sentences?
- Is it also the case in spontaneous speech that the nucleus is not (necessarily) perceived as more prominent than other items? It is generally accepted that spontaneous speech contains much more variation in the signalling of focus or information structure, so we might expect more nuclei to be perceived as particularly prominent simply because they have narrow focus. But are nuclei in unemphatic phrases with broad focus generally more prominent than other items in spontaneous speech?
- It has been argued in this thesis that the reduction in prominence level on the second lexical item, that is, the one following the onset, is related to its position in the utterance, but it could also be related to word class. Previous research (Widera *et al.* 1997, Streefkerk, Pols and ten Bosch 2001) has shown that verbs are generally less prominent than other lexical words, and because of the simple structure of the sentences in my experiment most of the lexical words in second position are verbs. The fact that the noun 'doctor' in the sentence 'Is Peter a doctor in Paris?' exhibits the same reduction as the verbs in the other sentences points to position as the more general explanation, but the issue calls for further research, using material with more variation in lexical and syntactic structure.

The acoustic correlates of prominence were largely unexplored and await further analysis. Among the key questions are the following:

- What are the acoustic correlates of the higher perceived prominence of the first and last lexical items?
- Is the strong – weak alternation in perceived prominence matched by variation in the acoustic parameters? If this is the case, then the alternation would appear to be planned by the speaker, and may be assumed to carry some (discourse related) meaning. If not, then it may be an artefact of the perception mechanism – that a certain prominence pattern is imposed on an otherwise regular prosodic structure.

- A first informal analysis of the acoustic data suggested a very direct link between perceived prominence and F_0, including the reduction of non-focal items which was visible in both prominence ratings and F_0 traces. Is the relation between perceived prominence and the (absolute) position of a stressed item in the utterance or the position relative to a focal accent matched by a similar relation between F_0 and position? And if F_0 is very directly related to this variation in perceived prominence, then how does variation in duration contribute to the overall picture?

While most of the questions about perceived prominence require new experiments with different material, most of the questions about the acoustic correlates can be addressed through the present material and build directly on the ratings and other results obtained in Tests 1–4.

Bibliography

Abercrombie, D. (1967): *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Adams, C. and R. R. Munro (1978): In search of the acoustic correlates of stress: fundamental frequency, amplitude and duration in the connected utterance of some native and non-native speakers of English. *Phonetica* 35: 125–156.

Armstrong, L. E. and I. C. Ward (1931): *Handbook of English intonation*. 2nd edn. Leipzig and Berlin: Teubner.

Beckman, M. E. and G. A. Elam (1997): *Guidelines for ToBI labelling*. Version 3.0, March 1997. Online publication: URL: http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf.

Berinstein, A. E. (1979): *A cross-linguistic study on the perception and production of stress*. UCLA Working Papers in Phonetics 47. Los Angeles: University of California, Los Angeles.

Bolinger, D. L. (1955): Intersections of stress and intonation. *Word* 11(2): 195–203.

Bolinger, D. L. (1958): A theory of pitch accent in English. *Word* 14(2-3): 109–149.

Bolinger, D. L. (1961): Contrastive accent and contrastive stress. *Language* 37(1): 83–96.

Bolinger, D. L. and L. J. Gerstman (1957): Disjuncture as a cue to constructs. *Word* 13(2): 246–255.

Brown, G., K. L. Currie and J. Kenworthy (1980): *Questions of intonation*. London: Croom Helm.

Buhmann, J., J. Caspers, V. J. van Heuven, H. Hoekstra, J.-P. Martens and M. Swerts (2002): Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. *Proceedings of LREC-2002*: 779–785. Las Palmas.

Cambier-Langeveld, T. (1999): The interaction between final lengthening and accentual lengthening: Dutch versus English. *Proceedings of the 14th International Congress of Phonetic Sciences*: 467–470. San Francisco.

Campbell, N. and M. Beckman (1997): Stress, prominence and spectral tilt. In A. Botinis, G. Kouroupetroglou and G. Carayannis (eds), *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Implications, Athens, Greece, September 18-20, 1997*, p. 67–70. Athens: ESCA and University of Athens.

Cohan, J. B. (2000): *The realization and function of focus in spoken English*. PhD dissertation. Austin: University of Texas at Austin.

Cohen, A. and J. 't Hart (1967): On the anatomy of intonation. *Lingua* 19: 177–192.

Coleman, H. O. (1914): Intonation and emphasis. *Miscellanea Phonetica* I: 6–26.

Couper-Kuhlen, E. (1986): *English prosody*. London: Edward Arnold.

Cruttenden, A. (1981): *The intonation of English sentences with special reference to adverbials*. Unpublished thesis. Manchester: University of Manchester.

Cruttenden, A. (1990): The origins of nucleus. *Journal of the International Phonetic Association* 20(1): 1–9.

Cruttenden, A. (1997): *Intonation*. 2nd edn. Cambridge: Cambridge University Press.

Crystal, D. (1969): *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Currie, K. L. (1978): Recent investigations in the field of intonation. *Work in Progress* 11: 63–77. Edinburgh: Department of Linguistics, Edinburgh University.

Currie, K. L. (1979): Further investigations in the field of intonation. *Work in Progress* 12: 20–29. Edinburgh: Department of Linguistics, Edinburgh University.

Davidsen-Nielsen, N. (1984): *Tonegangen i britisk engelsk*. Copenhagen: Albion.

Davidsen-Nielsen, N. (1994): *An outline of English pronunciation*. Odense: Odense University Press.

de Pijper, J. R. (1983): *Modelling British English intonation*. Dordrecht: Foris.

Eriksson, A., E. Grabe and H. Traunmüller (2002): Perception of syllable prominence by listeners with and without competence in the tested language. *Proceedings of Speech Prosody 2002*. Aix-en-Provence: Laboratoire Parole et Langage. URL: http://www.lpl.univ-aix.fr/sp2002/pdf/eriksson-grabe-traunmuller.pdf.

Eriksson, A., G. C. Thunberg and H. Traunmüller (2001): Syllable prominence: a matter of vocal effort, phonetic distinctness and top-down processing. *Proceedings of Eurospeech 2001 – Scandinavia*: 399–402. Aalborg: Center for Personkommunikation, Aalborg University.

Fant, G. and A. Kruckenberg (1989): Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR* 2/1989: 1–83.

Fant, G., A. Kruckenberg, J. Liljencrants and A. Botinis (2001): Prominence correlates. A study of Swedish. *Proceedings of Eurospeech 2001 – Scandinavia*: 657–660. Aalborg: Center for Personkommunikation, Aalborg University.

Ferguson, G. A. (1971): *Statistical analysis in psychology and education*. 3rd edn. New York: McGraw-Hill.

Fischer-Jørgensen, E. (1984): The acoustic manifestation of stress in Danish with particular reference to the reduction of stress in compounds. *Annual Report of the Institute of Phonetics, University of Copenhagen* 18: 45–161.

Fry, D. B. (1955): Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27: 765–768.

Fry, D. B. (1958a): Experiments in the perception of stress. *Language and Speech* 1: 126–152.

Fry, D. B. (1958b): The perception of stress. *Proceedings of the 8th International Congress of Linguistics, Oslo 1958*: 601–603. Oslo.

Fry, D. B. (1965): The dependence of stress judgments on vowel formant structure. *Proceedings of the 5th International Congress of Phonetic Sciences, Münster 1964*: 306–311.

Münster.

Gimson, A. C. (1956): The linguistic relevance of stress in English. *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 9: 143–149.

Gimson, A. C. (1989): *An introduction to the pronunciation of English*. 4th edn, revised by Susan Ramsaran. London: Edward Arnold.

Grønnum, N. (1992): *The groundworks of Danish intonation*. Copenhagen: Museum Tusculanum Press.

Grønnum, N. (1995): Superposition and subordination in intonation – a non-linear approach. *Proceedings of the 13th International Congress of Phonetic Sciences ICPhS 95* 13: 124–131. Stockholm: Department of Speech Communication and Music Acoustics, Royal Institute of Technology and Department of Linguistics, Stockholm University.

Grønnum, N. (2003): Dansk intonation. In A. Holmen, E. Glahn and H. Ruus (eds), *Veje til dansk – forskning i sprog og sprogtilegnelse*, p. 15–38. Copenhagen: Akademisk Forlag.

Gussenhoven, C., B. H. Repp, A. C. M. Rietveld, H. H. Rump and J. Terken (1997): The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America* 102(5): 3009–3022.

Gussenhoven, C. and A. C. M. Rietveld (1988): Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics* 16: 355–369.

Hakstian, A. R. and T. E. Whalen (1976): A k-sample significance test for independent alpha coefficients. *Psychometrica* 42(2): 219–231.

Handbook (1999): *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Heldner, M. (1996): Is an F0-rise a necessary or a sufficient cue to perceived focus in Swedish. In S. Werner (ed.), *Nordic Prosody: Proceedings of the VIIth Conference, Joensuu 1996*, p. 109–125. Frankfurt am Main: Peter Lang. URL: http://www.ling.umu.se/~heldner/papers/NordicProsodyVII.pdf.

Heldner, M. (2001a): *Focal accent - f0 movements and beyond*. (PHONUM 8). Umeå: Department of Philosophy and Linguistics, Umeå University.

Heldner, M. (2001b): Spectral emphasis as a perceptual cue to prominence. *TMH-QPSR* 2/2001: 51–57. URL: http://www.speech.kth.se/qpsr/tmh/01-2-051-057.pdf.

Heldner, M. (2001c): Spectral emphasis as and additional source of information in accent detection. *Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*: 57–60. Red Bank, NJ. URL: http://www.ling.umu.se/~heldner/papers/prosody2001.pdf.

Heldner, M. (2003): On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics* 31: 39–62. URL: http://www.speech.kth.se/~mattias/papers/JPHON%2703.pdf.

Heuft, B. and T. Portele (1996): Synthesizing prosody: a prominence-based approach. *Proceedings ICSLP 96: Fourth International Conference on Spoken Language Processing*: 1361–1364. IEEE Signal Processing, Acoustical Society of America.

Hirst, D. and A. Di Cristo (1998): A survey of intonation systems. In D. Hirst and A. Di Cristo (eds), *Intonation systems. A survey of twenty languages*, p. 1–44. Cambridge: Cambridge University Press.

Hitchcock, L. and S. Greenberg (2001): Vowel height is intimately associated with stress accent in spontaneous American English discourse. *Proceedings of Eurospeech 2001 – Scandinavia*: 79–82. Aalborg: Center for Personkommunikation, Aalborg University.

Huss, V. (1978): English word stress in the post-nuclear position. *Phonetica* 35: 86–105.

Jassem, W. (1952): Stress in modern English. *Bulletin de la Société Linguistique Polonaise* 11: 23–49.

Jassem, W. and D. Gibbon (1980): Re-defining English accent and stress. *Journal of the International Phonetic Association* 10(1): 2–16.

Jespersen, O. (1899): *Fonetik*. Copenhagen: Det Schubotheske Forlag.

Jones, D. (1909, 1967): *The pronunciation of English*. 4th edn. Cambridge: Cambridge University Press.

Jones, D. (1918, 1967): *An outline of English phonetics*. 9th edn. Cambridge: Heffer.

Jones, D. (1950, 1967): *The phoneme*. 3rd edn. Cambridge: Heffer.

Kingdon, R. (1958a): *The groundwork of English intonation*. London: Longman.

Kingdon, R. (1958b): *The groundwork of English stress*. London: Longman.

Knowles, G. (1987): *Patterns of spoken English*. London and New York: Longman.

Ladd, D. R. (1996): *Intonational phonology*. Cambridge: Cambridge University Press.

Lawlis, G. F. and E. Lu (1972): Judgment of counseling process: reliability, agreement, and error. *Psychological Bulletin* 78(1): 17–20.

Liberman, M. (1979): *The intonational system of English*. Outstanding Dissertations in Linguistics. New York and London: Garland Publishing.

Liberman, M. and J. Pierrehumbert (1984): Intonational invariance under changes in pitch range and length. In M. Aronoff and R. T. Oehrle (eds), *Language sound structure*, p. 157–233. Cambridge, MA: The MIT Press.

Livbjerg, I. and I. M. Mees (1997): *Practical English phonetics*. 2nd edn. Copenhagen: Det Schønbergske Forlag.

Mees, I. M. and B. Collins (2002): *Sound English*. 3rd edn. Copenhagen: Handelshøjskolens Forlag.

Nakatani, L. H. and C. H. Aston (1978): *Acoustic and linguistic factors in stress perception*. Unpublished manuscript.. Murray Hill, New Jersey: Bell Laboratories.

O'Connor, J. D. and G. F. Arnold (1961): *Intonation of colloquial English*. London: Longmans.

O'Connor, J. D. and G. F. Arnold (1973): *Intonation of colloquial English*. 2nd edn. London: Longman.

Palmer, H. E. (1922): *English intonation with systematic exercises*. Cambridge: Heffer.

Pierrehumbert, J. (1979): The perception of fundamental frequency declination. *Journal of the Acoustical Society of America* 66(2): 363–369.

Pierrehumbert, J. (1980): *The phonology and phonetics of English intonation*. Reproduced, November 1987. Bloomington, Indiana: Indiana University Linguistics Club.

Pike, K. L. (1943): *Phonetics*. Ann Arbor: University of Michigan Press.

Pitrelli, J. F., M. E. Beckman and J. Hirschberg (1994): Evaluation of prosodic transcription labelling reliability in the ToBI framework. *Proceedings of the International Conference on Spoken Language Processing (ICSLP), 1994*: 123–126. Yokohama, Japan.

Rietveld, A. C. M. and C. Gussenhoven (1985): On the relation between pitch excursion size and prominence. *Journal of Phonetics* 13: 299–308.

Rietveld, T. and R. van Hout (1993): *Statistical techniques for the study of language and language behaviour*. Berlin: Mouton de Gruyter.

Sautermeister, P. and R. Eklund (1997): Some observations on the influence of F_0 and duration to the perception of prominence by Swedish listeners. *Proceedings of Fonetik 97 (PHONUM 4)*: 121–124. Umeå: Department of Philosophy and Linguistics, Umeå University.

Scott, N. C. (July, 1939): An experiment on stress perception. *Le maître Phonétique* 67: 44–45.

Siegel, S. and N. J. Castellan, Jr. (1988): *Nonparametric statistics for the behavioral sciences*. 2nd edn. New York: McGraw-Hill.

Silipo, R. and S. Greenberg (1999): Automatic transcription of prosodic stress for spontaneous English discourse. *Proceedings of the 14th International Congress of Phonetic Sciences*: 2351–2354. San Francisco.

Silipo, R. and S. Greenberg (2000): Prosodic stress revisited: reassessing the role of fundamental frequency. *Proceedings of the NIST Speech Transcription Workshop*. URL: http://www.icsi.berkeley.edu/~steveng/2000/PDF/Prosody.pdf.

Silverman, K. E. A., M. E. Beckman, M. Ostendorf, C. W. Wightman, P. Price, J. Pierrehumbert and J. Hirschberg (1992): ToBI: A standard for labeling English prosody. *Proceedings of ICSLP 92* 2: 867–870. Alberta: Department of Linguistics, University of Alberta.

Sityaev, D. and J. House (2003): Phonetic and phonological correlates of broad, narrow and contrastive focus in English. *Proceedings of the 15th International Congress of Phonetic Sciences ICPhS*: 1819–1822. Barcelona, 2003.

Sluijter, A. M. C. (1995): *Phonetic correlates of stress and accent.* PhD thesis. Leiden: University of Leiden.

Sluijter, A. M. C., S. Shattuck-Hufnagel, K. N. Stevens and V. J. van Heuven (1995): Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English. *Proceedings of the 13th International Congress of Phonetic Sciences ICPhS 1995*. Stockholm.

Sluijter, A. M. C. and V. J. van Heuven (1993): Perceptual cues of linguistic stress: intensity revisited. *Proceedings of the ESCA Workshop on Prosody 1993*: 246–249. Lund: Department of Linguistics and Phonetics, Lund University. (Working Papers 41).

Sluijter, A. M. C. and V. J. van Heuven (1996): Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4): 2471–2485.

Sluijter, A. M. C., V. J. van Heuven and J. J. A. Pacilly (1997): Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America* 101(1): 503–513.

Streefkerk, B. M. (1997): Acoustical correlates of prominence: a design for research. *Proceedings* 21: 131–142. Amsterdam: Institute of Phonetic Sciences, University of Amsterdam.

Streefkerk, B. M. and L. C. W. Pols (1996): Prominent accent and pitch movements. *Proceedings* 20: 111–119. Amsterdam: Institute of Phonetic Sciences, University of Amsterdam.

Streefkerk, B. M., L. C. W. Pols and L. F. M. ten Bosch (1997): Prominence in read aloud sentences, as marked by listeners and classified automatically. *Proceedings* 21: 101–116. Amsterdam: Institute of Phonetic Sciences, University of Amsterdam.

Streefkerk, B. M., L. C. W. Pols and L. F. M. ten Bosch (1998): Automatic detection of prominence (as detected by listeners' judgements) in read aloud Dutch sentences. *Proceedings of ICSLP'98, Sydney, Australia, December 1998*: 683–686.

Streefkerk, B. M., L. C. W. Pols and L. F. M. ten Bosch (1999): Towards finding optimal features of perceived prominence. *Proceedings of the 14th International Congress of Phonetic Sciences*: 1769–1772. San Francisco.

Streefkerk, B. M., L. C. W. Pols and L. F. M. ten Bosch (2001): Acoustical and lexical/syntactic features to predict prominence. *Proceedings* 24: 155–166. Amsterdam: Institute of Phonetic Sciences, University of Amsterdam.

Terken, J. (1991): Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America* 89(4): 1768–1776.

't Hart, J. and A. Cohen (1973): Intonation by rule, a perceptual quest. *Journal of Phonetics* 1: 309–327.

't Hart, J., R. Collier and A. Cohen (1991): *A perceptual study of intonation*. Cambridge: Cambridge University Press.

Thorsen, N. (1978): An acoustical investigation of Danish intonation. *Journal of Phonetics* 6: 151–175.

Thorsen, N. (1980): Neutral stress, emphatic stress, and sentence intonation in Advanced Standard Copenhagen Danish. *Annual Report of the Institute of Phonetics, University of Copenhagen* 14: 121–205.

Thorsen, N. (1982): On the variability in F0 patterning and the function of F0 timing in languages where pitch cues stress. *Phonetica* 39: 302–316.

Tinsley, H. E. A. and D. J. Weiss (1975): Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology* 22(4): 358–376.

Turk, A. E. and J. R. Sawusch (1997): The domain of accentual lengthening in American English. *Journal of Phonetics* 25(1): 25–41.

Wagner, P. (1999): The synthesis of German contrastive focus. *Proceedings of the 14th International Congress of Phonetic Sciences*: 1529–1532. San Francisco.

Wagner, P. and E. Fischenbeck (2002): Stress perception and production in German stress clash environments. *Proceedings of Speech Prosody 2002*. Aix-en-Provence:

Laboratoire Parole et Langage. URL: http://www.lpl.univ-aix.fr/sp2002/pdf/wagner.pdf.

Wagner, P. and T. Portele (1999): Two dimensions of prominence. *Proceedings of the ESCA Workshop on Dialogue and Prosody 1999*. Eindhoven.

Widera, C., T. Portele and M. Wolters (1997): Prediction of word prominence. *Proceedings of Eurospeech'97*: 999–1002. Rhodes.

Wightman, C. W. (1993): Perception of multiple levels of prominence in spontaneous speech (abstract). *ASA 126th Meeting Denver 1993*. URL: http://www.auditory.org/asamtgs/asa93dnv/5aSP/5aSP10.html.

Wightman, C. W. (2002): ToBI or not ToBI?. *Proceedings of Speech Prosody 2002*. Aix-en-Provence: Laboratoire Parole et Langage. URL: http://www.lpl.univ-aix.fr/sp2002/pdf/wightman.pdf.

Wightman, C. W. and M. Ostendorf (1994): Automatic labeling of prosodic patterns. *IEEE Transactions on Speech and Audio Processing* 4(2): 469–481.

Willems, N. (1982): *English intonation from a Dutch point of view*. Dordrecht: Foris.

Winer, B. J., D. R. Brown and K. M. Michels (1971): *Statistical principles in experimental design*. New York: McGraw-Hill.

Appendix contents

APPENDIX A

Information about experimental procedures

A.1 Information about speakers

All the speakers in the experiment lived in Edinburgh at the time of the recordings and had done so for considerable, but variable, lengths of time. They all described their accents as RP or southern English – some with slight modifications that appear from the descriptions below which in addition to their age and sex present the comment they (or others) made about the accent or (linguistic) background.

1F Speaker 1F is female and was 29 at the time of the recordings; born to American parents, but grew up in Brighton, England. Had lived in Edinburgh for five years and commented that that might have influenced her accent, but stated that '[she] was deliberately speaking RP with [me]'.

2F 2F is female, 40 years of age and from East Sussex. Her own written comments about her pronunciation were: 'Not really "pure" RP in actual conversation (though it was in the recording studio!). There's a little influence of the "South London" accent (definitely low prestige).'

3F Female, age 30. Born in Bristol, but lived in Sheffield through her secondary school years (age 8 - 18). Both her parents are native speakers of RP, and her accent was described by one of her (phonetician) colleagues as 'definitely "young" RP'.

4M 4M is male and at 59 the oldest of the six speakers. He was born in Wales to an English-speaking father and Welsh-speaking mother and was brought up bilingually. Moved to London at age six where he went to school; later university studies and seven years of EFL teaching in Cambridge, before moving to Edinburgh at age 30. He comments on his accent that '[he] think[s] there is a residual Welshness under [his] RP'.

5M Speaker 5M is male and 37 of age. He comments that he is from 'the South of England'.

6M Speaker 6M is male and was 43 at the time of the recording. He was born in London, but spent his early childhood 'all over the world'. Started to speak RP at age 7 to 9 and his accent was described by a colleague of his as 'quite RP'.

Although the speakers are not all native monolingual speakers of either traditional public school RP or a more modern standard English variety such as 'non-regional pronunciation' (Collins and Mees 2003), they do represent, at least to some extent,

Southern Standard British English. Readers who wish to evaluate this claim for themselves can hear samples of their speech at the webpage accompanying this thesis.

A.2 Recording setup and calibration

Recording setup

The recording setup included two microphones. In order to measure intensity confidently it is necessary that the distance between speaker and microphone is constant and that the recording level is calibrated, in case cross-speaker comparisons are required. One microphone was of the 'clip-on' type which was fastened to a headband on a 15 cm long extension built of Meccano® (a metal construction toy). In general this worked fairly well, but it was not entirely comfortable to wear, and on one of the speakers it could not be fastened securely. There are also some instances of microphone noise caused by aspiration bursts on aspirated stops, if the microphone had been placed too low, since it was always placed directly in front of the mouth. The distance from the mouth was approximately 13 cm for all the speakers. To counteract some of the problems with this method a 'shotgun' type directional microphone was also used, placed approximately 1 meter from the speaker. There are two reasons for this placement: first of all the setup is less vulnerable to variations in the distance from speaker to microphone when this distance is long (simply minimising the per cent variation), and secondly there are several studies which have shown a considerable effect of microphone placement on the measurement of intensity (Ludvigsen 1971, Ludvigsen and Thorsen 1971, Ludvigsen 1979), and the 1979 study recommends a long distance from the mouth (for example 1 meter). This seems particularly important in connection with measurements of spectral balance, or slope:

> If recordings are used to study the intensity relation between different frequency components of a sound, the microphone position will be important. If the recordings, furthermore, are used to estimate e.g. the slope of the glottis spectrum by means of inverse filtering, appreciable differences may appear due to different microphone positions (Ludvigsen 1979:186).

By using two different microphones it was possible to compare measurements and thereby get the best of both worlds.

Both microphone signals were fed through a studio mixer and recorded on to separate tracks of a DAT recorder. The digitising of the microphone signals was performed by the DAT recorder, and the digital output from the DAT was connected to a digital audio interface which converted the signal to 16 kHz (from the DAT's 48 kHz) and stored it on a Sun workstation through a digital link. For one of the speakers, (3F), this last process was omitted and the digitising had to be done on a Sun UltraSPARC through the internal sound card of this machine, using the ESPS/Ensig® program. This signal was also stored in 16 bit, 16 kHz format with the two microphone signals on separate tracks.

Calibration

In order to be able to compare intensity level one needs a basis for comparison. Therefore the recording level was calibrated before (and after) each speaker session (a complete recording of one speaker) using a sound pressure level meter, and not altered during this session. A 1 kHz reference tone was recorded at a fixed distance from the clip-on microphone and the level adjusted to about 72 dB. Using this method it is possible to evaluate intensity levels relative to an absolute measure (the reference, or calibration, tone), both for one speaker and across speakers.

Equipment

The clip-on microphone was a Sennheiser MKE 2 with a Sennheiser K6 powering module. The 'shotgun' microphone was an AKG Blue Line (CK98 capsule plus SE 300 B powering module). Both signals were fed through a Soundtrac 200B Studio Console and recorded on to separate tracks of a Sony PCM2700A DAT recorder.
The digital link was a Townsend Datlink II digital audio interface which was connected to a Sun IPC workstation.

Post-recording correction

One or two errors occurred during the recording process which had to be corrected at a later stage: For speaker 4M both signals had been recorded on to one track. This was resolved by subtracting the track with one microphone signal from the one containing both. Furthermore, because of a fault in the studio setup, the two microphone channels were not always recorded to the same track, that is, sometimes the clip-on microphone was recorded to track 0, sometimes to track 1. This could even vary within one speaker session, between part-sessions. It was necessary to determine which was which auditorily (which was quite easy since they sounded very different).

A.3 Instructions to the raters in the listening experiment

A.3.1 Test 1 – instructions to the Danish raters

Tryk og prominens

Markering af prominens og prosodiske grænser i engelske sætninger

Tak fordi I har sagt ja til at være med til dette lytteeksperiment. Opgaven går ud på at markere prominens og prosodiske grænser i en række korte engelske sætninger efter følgende principper:

Prominens:

Der opereres med 3 niveauer af prominens:

- Kraftigt tryk, markeres med to streger oppe, fx ’ ’Peter did it.
- Almindeligt tryk, markeres med en enkelt streg oppe: fx ’Peter
- Svagere tryk, markeres med enkelt streg nede, fx ,Peter ’ ’didn’t do it.

Svagtryk markeres ikke

Prosodiske grænser:

Der skelnes mellem 2 niveauer af prosodiske grænser:

- Stærk prosodisk grænse markeres med //.
- Svagere prosodisk grænse markeres med /.

Bemærk at eftersom testen består af korte sætninger, kan der ikke forventes mange prosodiske grænser.

Testen er delt op i 4 dele på hver ca. 45 sætninger. Lyt venligst til sætningerne i den rækkefølge de optræder i de 4 dele af testen.

Jeg anbefaler at man benytter hovedtelefoner ved aflytningen. Alternativt et par gode højttalere tilsluttet computeren. Anfør venligst hvilken opstilling der blev benyttet.

I må lytte til hver enkelt sætning lige så mange (eller så få) gange som er nødvendigt for at kunne lave markeringerne. De delsætninger som står i parentes, skal ikke trykmarkeres.

God fornøjelse!

Første del af testen

Anden del af testen

Tredje del af testen

Fjerde del af testen

A.3.2 Test 2 – instructions to the English raters

Listening experiment

Marking stress in English utterances

Thank you for agreeing to participate in this listening experiment. The task is to mark stress in a number of short English utterances according to the following conventions:

ˌˌ indicates extra strong stress.
ˌ indicates (normal) full stress.
ˌ indicates reduced stress.

You can mark three degrees of stress, or fewer, as you deem appropriate.

(Completely unstressed words/syllables are not marked explicitly.)

The test is in 4 parts, each consisting of about 45 utterances. Please listen to the utterances in the order in which they occur in the 4 parts of the test.

I recommend the use of headphones; alternatively a pair of good speakers connected to your computer. Please indicate which option you have used.

You can listen to each utterance as many (or as few) times as necessary. The bracketed parts of the sentences need not be marked. You should have received the answer sheets separately. If not they are available as a PDF file here.

Best wishes, and enjoy!

Christian Jensen

First part of the test

Second part of the test

Third part of the test

Fourth part of the test

A.3.3 Test 3 – instructions to the 'British school' raters

Stress and accent

Marking stress and accent in English utterances

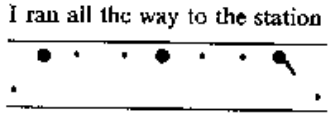
Thank you for agreeing to participate in this listening experiment. The task is to mark stress and accent in a series of short English utterances according to the following principles (from Cruttenden 1997:18+44):

3 degrees of stress/accent (+ unstressed) are distinguished.

- Primary stress/accent, involving the principal pitch prominence in the intonation-group, i.e. the nucleus.
  Indicate by the raised number "1".
- Secondary stress/accent, involving a subsidiary pitch prominence in an intonation-group, i.e. a non-nuclear pitch accent.
  Indicate by the raised number "2".
- Tertiary stress, involving a prominence produced principally by length and/or loudness.
  Indicate by the raised number "3".

Unstressed syllables are not marked.

The following example sentence from Cruttenden (p. 18)



would thus look like this:

```
   2          3          1
I ran all the way to the station
```

If an utterance consists of more than one intonation-group the boundary can be marked by a "/".

> NB! If you feel more comfortable using a regular 'tonetic stress-mark' system, you are welcome to do so. I do not need information about the *type* of nucleus, so please do not concentrate too much energy on that. Using tonetic stress-marks the sentence could look like this:
>
> ```
> I ˈran all the *way to the ˈstation
> ```

The test is divided up into 4 parts, each consisting of about 45 utterances. Please listen to the utterances in the order in which they occur in the 4 parts of the test.

I recommend the use of headphones; alternatively a pair of good speakers connected to your computer. Please indicate which option you have used.

You can listen to each utterance as many (or as few) times as necessary. The parts of the utterances which are bracketed should not be marked. You should have received the answer sheets separately.

If not it is available as a PDF file here.

Best wishes, and enjoy!

Christian Jensen

First part of the test

Second part of the test

Third part of the test

Fourth part of the test

APPENDIX B

Data listings

B.4 Segment durations and ratios – individual speakers

Segment durations and ratios – Speaker 1F											
Sent	*Lex*	*Seg*	*n*	*f1*	*f2*	*f3*	*f4*	*f1/n*	*f2/n*	*f3/n*	*f4/n*
ps	1	ɔːl	194	237	158			**1.22**	0.82		
ps	2	ɪŋ	254	219	263			0.86	**1.04**		
pc	1	ɑː	109	122	101			**1.11**	0.92		
pc	2	æn	143	158	138			1.10	**0.96**		
bsa	1	ɪl	192	203	152	141		**1.06**	0.79	0.73	
bsa	2	ʌ	170	162	189	155		0.95	**1.16**	0.91	
bsa	3	æ	199	165	170	208		0.83	0.86	**1.05**	
css	1	ʊ	75	71	66	60		**0.95**	0.89	0.80	
css	2	e	50	56	66	47		1.12	**1.32**	0.95	
css	3	uː	183	190	194	204		1.04	1.06	**1.11**	
jkft	1	eɪn	242	264	199	183	193	**1.09**	0.82	0.76	0.80
jkft	2	ɪ	58	56	62	50	46	0.96	**1.07**	0.85	0.80
jkft	3	ræŋ	124	108	111	137	110	0.87	0.89	**1.10**	0.89
jkft	4	en	137	129	133	136	138	0.94	0.97	0.99	**1.01**
sepc	1	iː	83	90	66	66	62	**1.08**	0.79	0.80	0.75
sepc	2	æ	85	79	96	76	79	0.93	**1.13**	0.89	0.93
sepc	3	eɪ	118	115	113	103	110	0.98	0.96	**0.87**	0.93
sepc	4	eə	138	139	132	131	147	1.01	0.96	0.95	**1.06**
dsi	1	ɑːlɪ	234	256				**1.09**			
dsi	2	ɪ	49	53				1.08			

Segment durations and ratios – Speaker 2F

Sent	*Lex*	*Seg*	*n*	*f1*	*f2*	*f3*	*f4*	*f1/n*	*f2/n*	*f3/n*	*f4/n*
ps	1	ɔːl	209	225	164			**1.08**	0.78		
ps	2	ɪŋ	312	275	326			0.88	**1.04**		
pc	1	ɑː	118	139	101			**1.18**	0.86		
pc	2	æn	127	141	127			1.10	**1.00**		
bsa	1	ɪl	200	225	177	156		**1.12**	0.88	0.78	
bsa	2	rʌ	116	109	114	109		0.94	**0.98**	0.94	
bsa	3	æ	175	179	180	187		1.02	1.03	**1.07**	
css	1	ʊ	72	74	56	61		**1.03**	0.78	0.84	
css	2	e	75	78	83	68		1.04	**1.1**1	0.91	
css	3	uː	122	117	114	135		0.96	0.93	**1.11**	
jkft	1	eɪn	244	267			207	**1.09**			0.85
jkft	2	ɪ	53	58			46	1.09			0.87
jkft	3	ræŋ	163	149			149	0.91			0.91
jkft	4	en	131	129			148	0.98			**1.13**
sepc	1	iː	112	109		109	100	**0.97**		0.97	0.89
sepc	2	æ	89	87		89	86	0.97		0.99	0.97
sepc	3	eɪ	93	98		103	97	1.06		**1.11**	1.04
sepc	4	eə	140	122		131	152	0.87		0.94	**1.08**
pdp	1	iː	81	80				**0.99**			
pdp	2	ɒ	81	82				1.01			
pdp	3	ærɪ	229	208				0.91			
dsi	1	ɑː	145	144				**0.99**			
dsi	2	ɪ	80	70				0.87			

Segment durations and ratios – Speaker 3F

Sent	*Lex*	*Seg*	*n*	*f1*	*f2*	*f3*	*f4*	*f1/n*	*f2/n*	*f3/n*	*f4/n*
ps	1	ɔːl	244	300	173			**1.23**	0.71		
ps	2	ɪŋ	304	299	410			0.98	**1.35**		
pc	1	ɑː	138	160	117			**1.16**	0.85		
pc	2	æn	125	134	164			1.07	**1.31**		
bsa	1	ɪl	222	256	146	139		**1.15**	0.66	0.63	
bsa	2	rʌ	111	113	137	107		1.02	**1.24**	0.97	
bsa	3	æ	190	156	164	223		0.82	0.86	**1.17**	
css	1	ʊ	111	98	67	81		**0.88**	0.60	0.73	
css	2	e	55	58	77	57		1.05	**1.38**	1.03	
css	3	uː	150	127	135	173		0.85	0.90	**1.15**	
jkft	1	eɪn	247	261	167	168	173	**1.05**	0.68	0.68	0.70
jkft	2	ɪ	50	47	65	52	46	0.95	**1.29**	1.05	0.93
jkft	3	ræŋ	165	133	146	175	140	0.80	0.88	**1.06**	0.85
jkft	4	en	127	126	131	134	151	0.99	1.04	1.06	**1.19**
sepc	1	iː	112	148	90	84	96	**1.33**	0.80	0.76	0.86
sepc	2	æ	85	82	129	87	90	0.96	**1.53**	1.02	1.07
sepc	3	eɪ	108	99	107	114	102	0.92	0.99	**1.06**	0.94
sepc	4	eə	125	113	113	114	145	0.90	0.90	0.91	**1.16**
pdp	1	iː	98	87				**0.89**			
pdp	2	ɒ	83	92				1.11			
pdp	3	ærɪ	216	197				0.91			

Segment durations and ratios – Speaker 4M											
Sent	*Lex*	*Seg*	*n*	*f1*	*f2*	*f3*	*f4*	*f1/n*	*f2/n*	*f3/n*	*f4/n*
ps	1	ɔːl	236	229	154			**0.97**	0.65		
ps	2	ɪŋ	249	194	273			0.78	**1.10**		
pc	1	ɑː		108	90						
pc	2	æn		133	151						
bsa	1	ɪl	180	162	150	132		**0.90**	0.83	0.73	
bsa	2	rʌ	101	98	98	96		0.97	**0.97**	0.95	
bsa	3	æ	158	124	141	154		0.78	0.89	**0.97**	
css	1	ʊ	49	61	47	52		**1.25**	0.95	1.07	
css	2	e	62	61	64	57		0.97	**1.03**	0.92	
css	3	uː	96	93	96	116		0.97	1.00	**1.21**	
jkft	1	eɪn	234	307	194	171	192	**1.31**	0.83	0.73	0.82
jkft	2	ɪ	59	69	51	54	54	1.17	**0.88**	0.93	0.92
jkft	3	ræŋ	157		139	168	131		0.89	**1.07**	0.83
jkft	4	en	126		121	117	128		0.96	0.93	**1.01**
sepc	1	iː	87	89		71		**1.02**		0.81	
sepc	2	æ	78	73		71		0.93		0.91	
sepc	3	eɪ	93	99		94		1.06		1.01	
sepc	4	eə	101	97		98		0.95		**0.96**	
pdp	1	iː	91	82				**0.90**			
pdp	2	ɒ	86	85				0.99			
pdp	3	ærɪ	225	178				0.79			

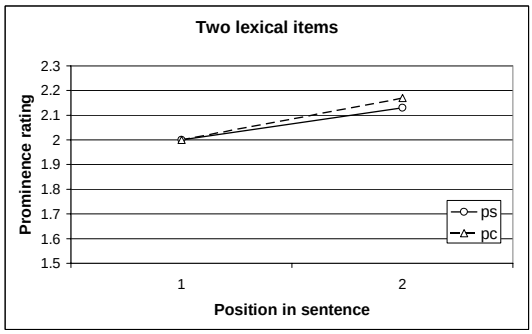
Segment durations and ratios – Speaker 5M											
Sent	*Lex*	*Seg*	*n*	*f1*	*f2*	*f3*	*f4*	*f1/n*	*f2/n*	*f3/n*	*f4/n*
ps	1	ɔːl	154	245	145			**1.59**	0.94		
ps	2	ɪŋ	292	282	334			0.96	**1.14**		
pc	1	ɑː	121	127	114			**1.05**	0.94		
pc	2	æn	133	138	141			1.04	**1.06**		
bsa	1	ɪl	171	214	151			**1.25**	0.89		
bsa	2	rʌ	98	87	105			0.89	**1.07**		
bsa	3	æ	214	192	187			0.90	0.87		
css	1	ʊ	59	83	47	56		**1.41**	0.79	0.95	
css	2	e	63	62	76	62		0.99	**1.21**	0.99	
css	3	uː	103	93	92	99		0.90	0.89	**0.96**	
jkft	1	eɪn	200	255	195	172	172	**1.27**	0.97	0.86	0.86
jkft	2	ɪ	61	60	65	52	44	0.98	**1.06**	0.85	0.73
jkft	3	ræŋ	125	115	112	156	106	0.92	0.90	**1.25**	0.85
jkft	4	en	132	144	150	132	138	1.09	1.14	1.00	**1.04**
sepc	1	iː	77	97	72	66	68	**1.26**	0.94	0.86	0.89
sepc	2	æ	88	86	110	82	82	0.98	**1.25**	0.93	0.93
sepc	3	eɪ	107	106	114	96	98	0.99	1.07	**0.89**	0.92
sepc	4	eə	108	100	101	102	110	0.92	0.94	0.94	**1.02**
pdp	1	iː		71							
pdp	2	ɒ		82							
pdp	3	ærɪ		196							

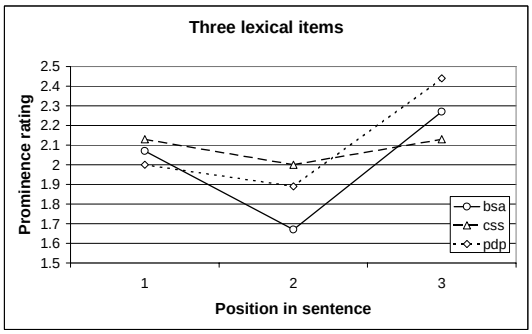
Segment durations and ratios – Speaker 6M											
Sent	*Lex*	*Seg*	*n*	*f1*	*f2*	*f3*	*f4*	*f1/n*	*f2/n*	*f3/n*	*f4/n*
ps	1	ɔːl	216	299	200			**1.38**	0.93		
ps	2	ɪŋ	366	338	447			0.92	**1.22**		
pc	1	ɑː	151	159	137			**1.05**	0.90		
pc	2	æn	166	160	163			0.96	**0.98**		
bsa	1	ɪl	157	244	120	127		**1.55**	0.76	0.81	
bsa	2	rʌ	98	107	113	87		1.10	**1.16**	0.89	
bsa	3	æ	247	230	220	295		0.93	0.89	**1.19**	
css	1	ʊ	82	97	64	75		**1.18**	0.78	0.91	
css	2	e	70	68	80	64		0.97	**1.14**	0.91	
css	3	uː	133	118	116	119		0.89	0.88	**0.90**	
jkft	1	eɪn	226	337			181	**1.49**			0.80
jkft	2	ɪ	66	60			63	0.91			0.96
jkft	3	ræŋ	168	185			162	1.10			0.96
jkft	4	en	142	141			154	0.99			**1.09**
sepc	1	iː	100		81		76		0.81		0.76
sepc	2	æ	101		130		97		**1.29**		0.97
sepc	3	eɪ	119		133		110		1.12		0.92
sepc	4	eə	163		158		155		0.97		**0.95**
pdp	1	iː	81	90				**1.11**			
pdp	2	ɒ	82	71				0.86			
pdp	3	ærɪ	216	221				1.03			
dsi	1	ɑː	128	152				**1.19**			
dsi	2	ɪ	69	75				1.07			

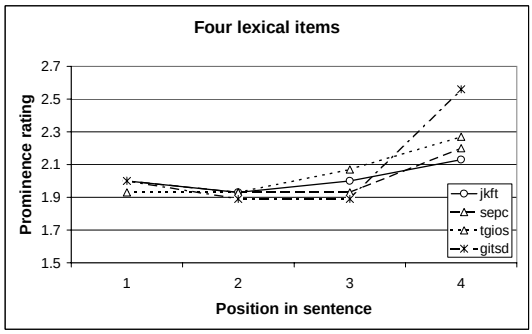
B.5 Prominence ratings – three Danish groups of raters

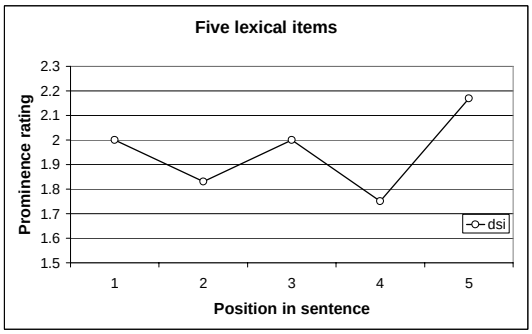
B.5.1 Neutral, context-free utterances

Group 1:

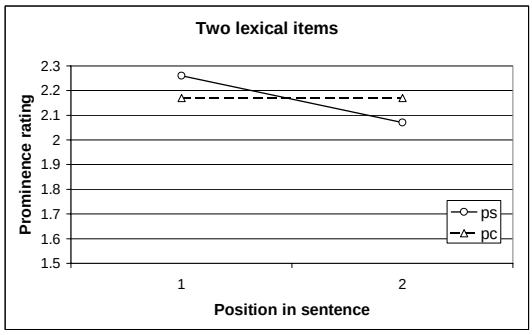


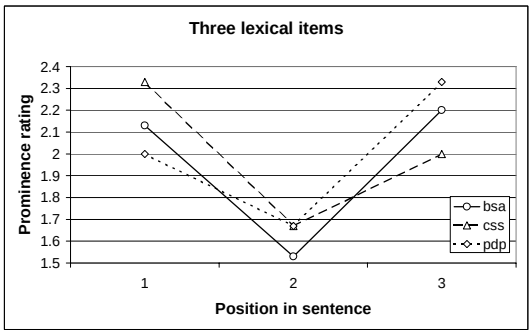


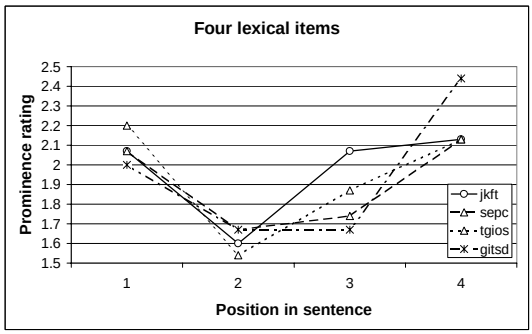


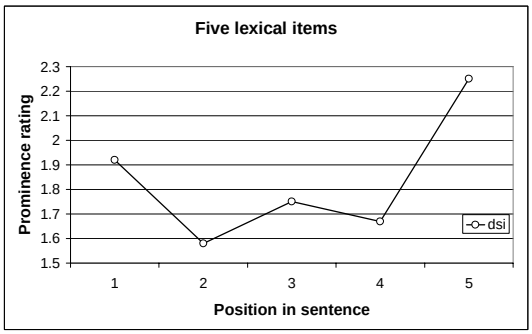


Group 2:









Group 3:

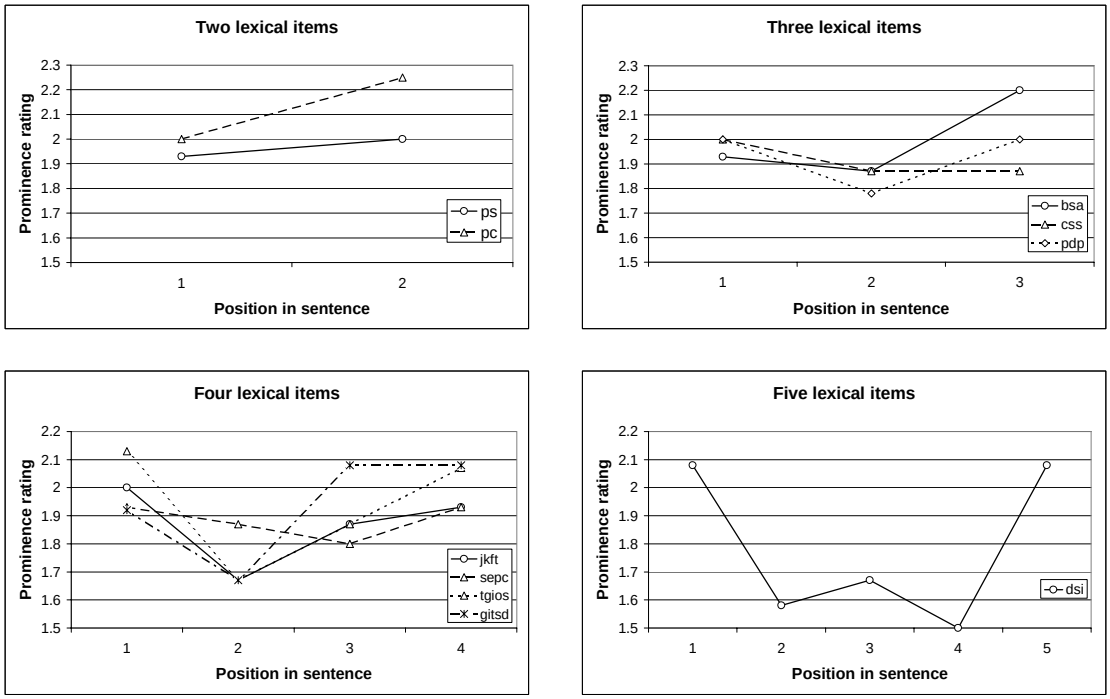


Figure B.1. Prominence ratings for the three groups of Danish raters – neutral utterances. See Section 3.4.1 for a description of the three groups.

B.5.2 Utterances with marked information structure

Group 1:

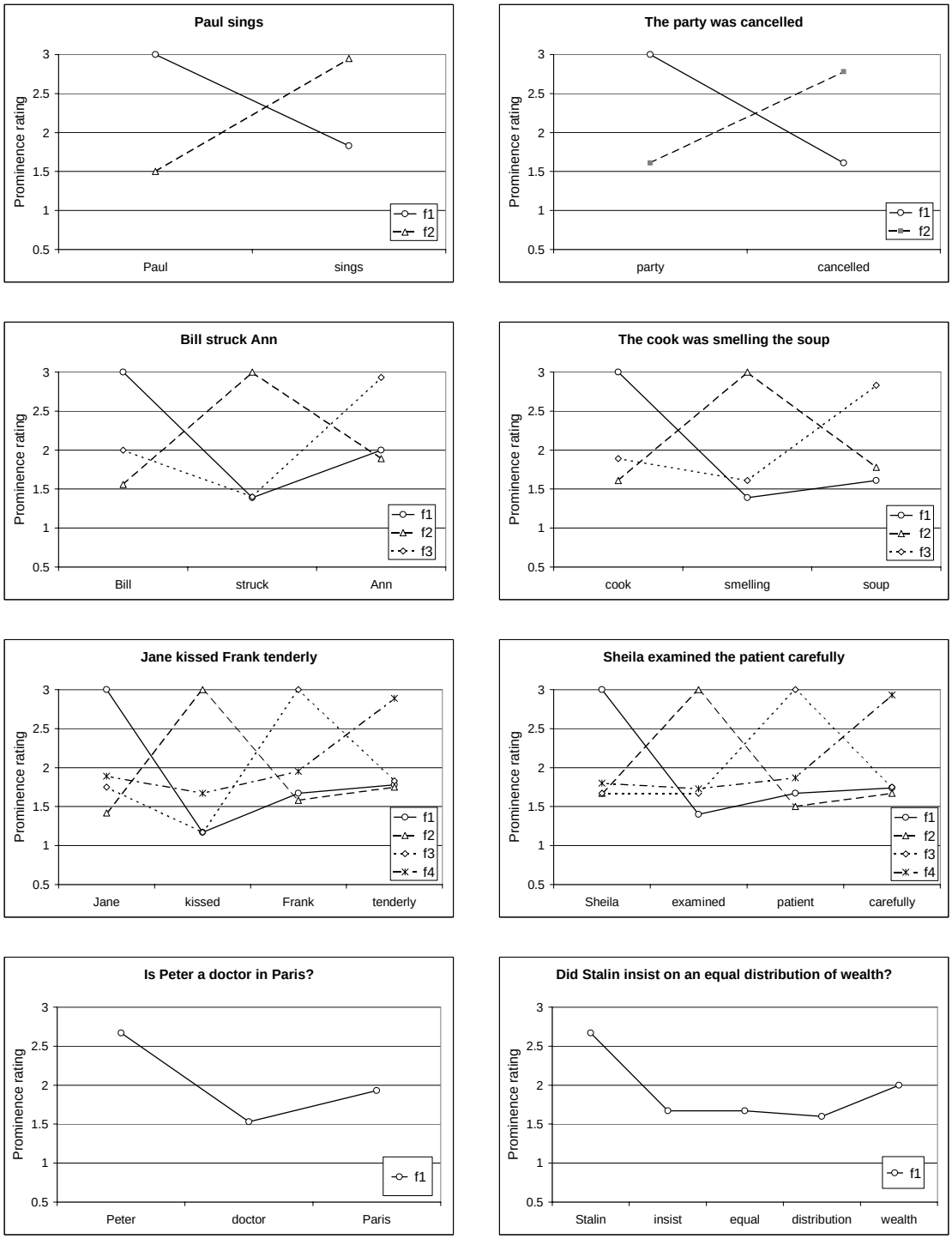


Figure B.2. Prominence ratings for Group 1 – utterances with marked information structure.

Group 2:

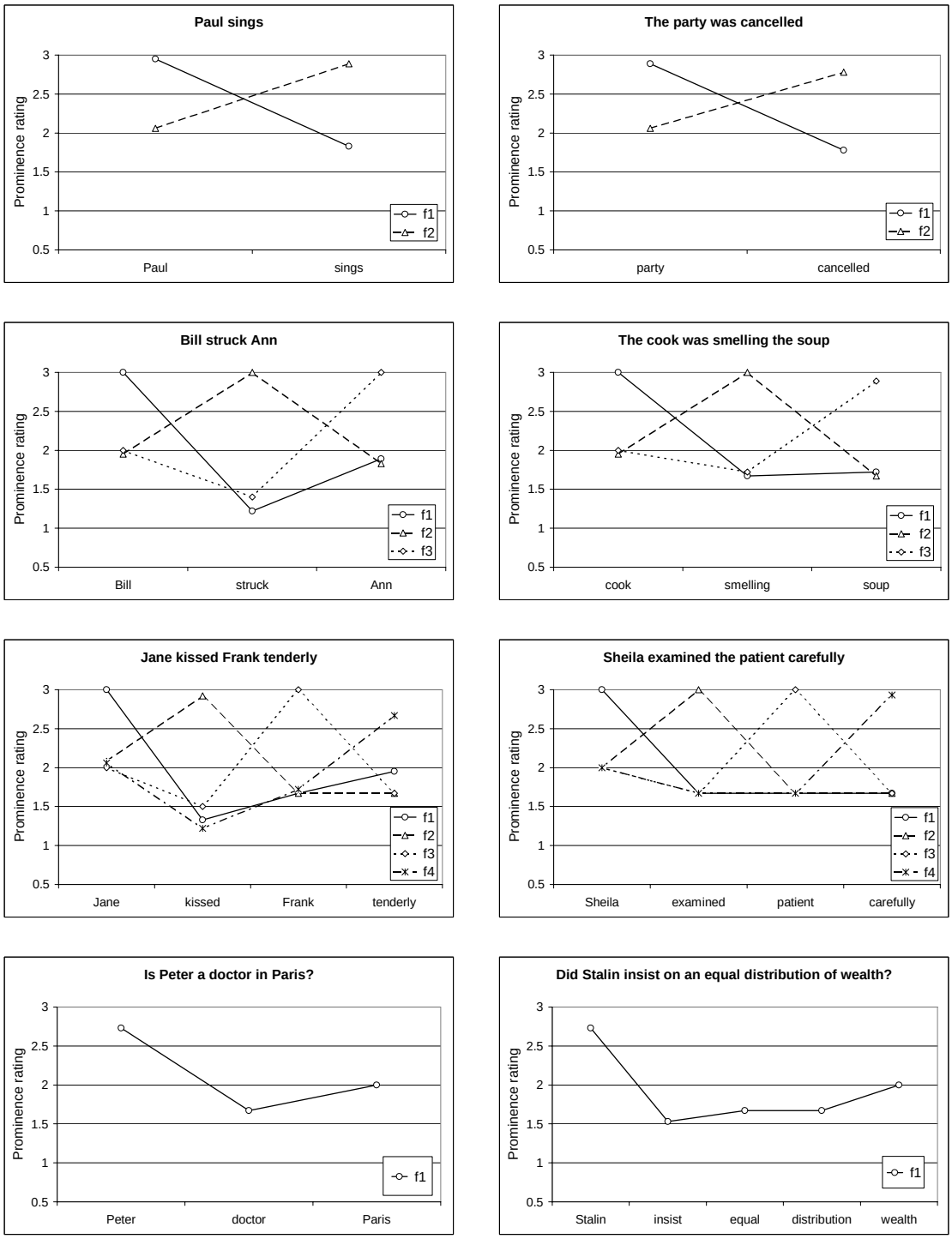


Figure B.3. Prominence ratings for Group 2 – utterances with marked information structure.

Group 3:

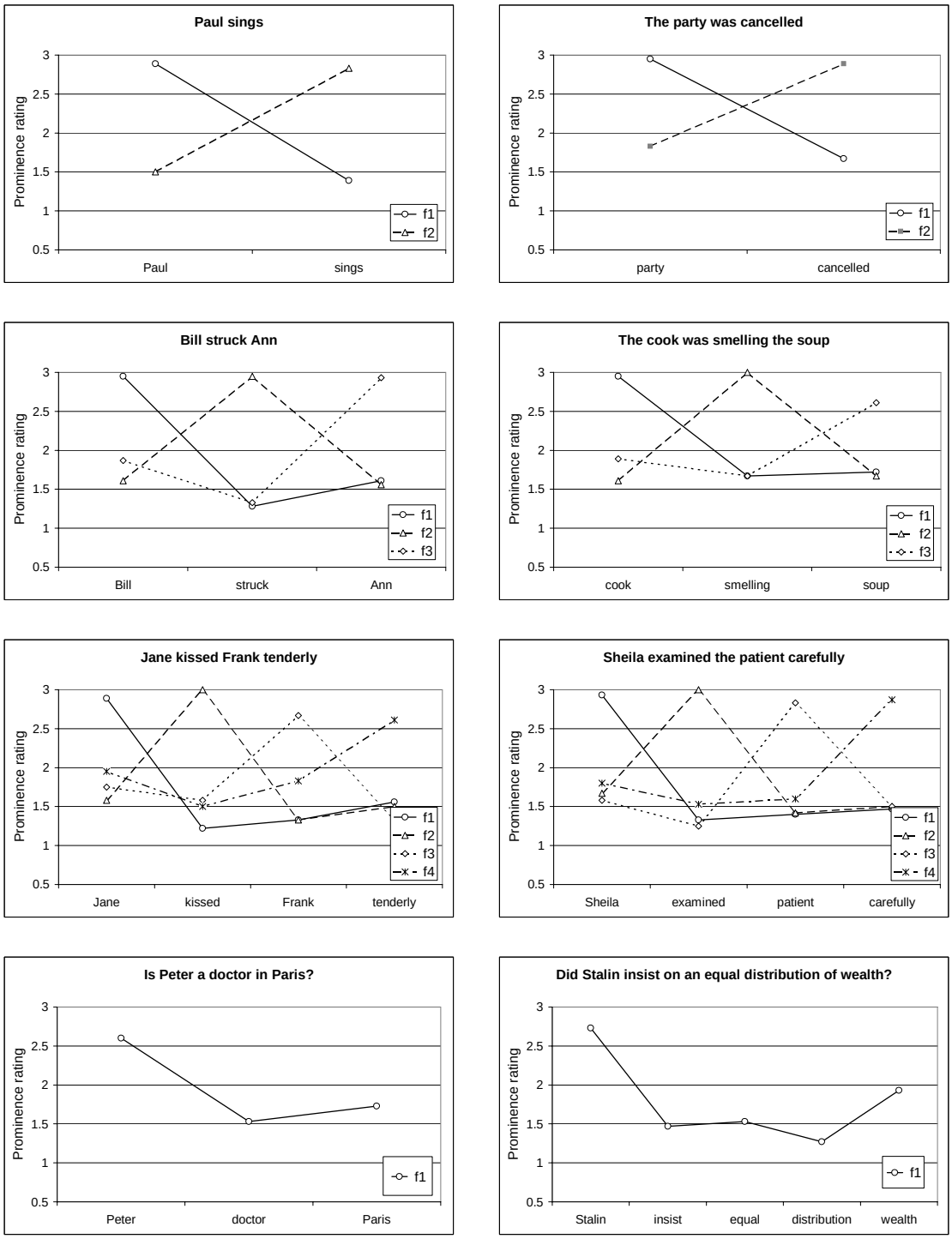
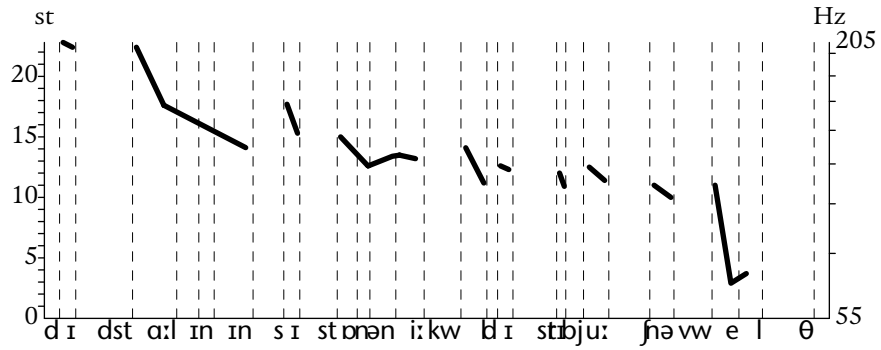


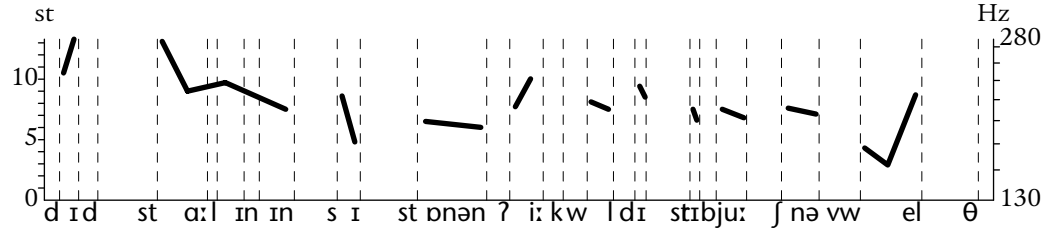
Figure B.4. Prominence ratings for Group 3 – utterances with marked information structure.

B.6 High preheads

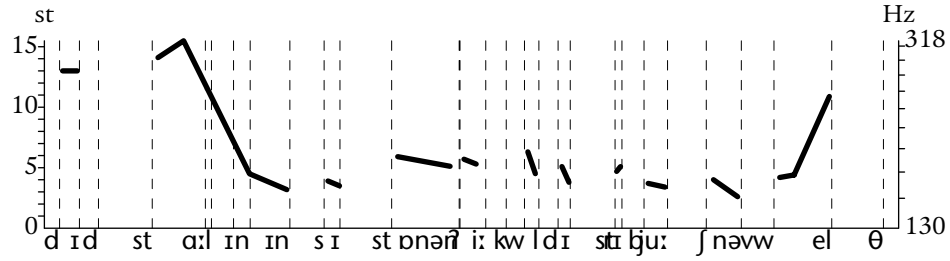
The following images are F_0 traces of the remaining three utterances treated in Section 6.9, but not shown in Figure 6.6.



Sent *dsi n* Spk 4M



Sent *dsi n* Spk 2F



Sent *dsi f1* Spk 2F

Bibliography (Appendix)

Collins, B. and I. M. Mees (2003): *Practical phonetics and phonology*. London: Routledge.

Ludvigsen, C. (1971): Energy measurements of speech sounds. *Annual Report of the Institute of Phonetics, University of Copenhagen* 5: 153–162.

Ludvigsen, C. (1979): Influence of microphone position in the recording of speech signals. *Annual Report of the Institute of Phonetics, University of Copenhagen* 13: 171–187.

Ludvigsen, C. and N. Thorsen (1971): Comparison of sound pressure level and loudness (GF) measurements on speech sounds. *Annual Report of the Institute of Phonetics, University of Copenhagen* 5: 163–174.